Section VI:
DSA Workshop Exhibits

Exhibit 204

From: "CNECT-DSA-REGISTRY@ec.europa.eu" <CNECT-DSA-REGISTRY@ec.europa.eu>
Date: Wednesday, 16 April 2025 at 08:13
Cc: "CNECT-DSA-REGISTRY@ec.europa.eu" <CNECT-DSA-REGISTRY@ec.europa.eu>, "CNECT-DIGITAL-SERVICES@ec.europa.eu" <CNECT-DIGITAL-SERVICES@ec.europa.eu>, GRAEF Eike <[redacted]@ec.europa.eu>, ZARRA Antonella <[redacted]@ec.europa.eu>
Subject: [EXTERNAL] Information about the upcoming DSA multi-stakeholder workshop on systemic risks

CAUTION: This email originated from outside of the organization. Do not click links or open attachments unless you can confirm the sender and know the content is safe.

Dear Sir/Madam,

In view of the upcoming DSA Multi-stakeholder workshop on Systemic Risks, which will take place on 7 May 2025, we want to give you some additional details on the format of the workshop. The workshop will take place at the European Commission – CCAB Centre de Conference Albert Borschette, Rue Froissart 36, 1049 Brussels.

After an introduction in the plenary session, participants will be divided into four tracks, happening in parallel, focusing on different systemic risks categories: Track 1 - Dissemination of illegal content, Track 2 - Disinformation and elections, Track 3 - Protection of minors and mental health, and Track 4 - Consumer protection.
Overall, the objectives of the workshop are to foster collaboration on risk assessments, help participants learn about the providers' different approaches to the identification, assessment, and mitigation of systemic risks, and gather perspectives from Digital Services Coordinators (DSCs), civil society organisations (CSO), researchers, and providers of VLOPs and VLOSEs.
Each session of the four tracks will be structured around different scenarios that depict different risks. These scenarios will guide the participants through a proactive discussion on approaches to assess and mitigate certain risks. The scenarios as well as the guiding questions will be shared in advance for the participants to be prepared. After an introduction into the specific risk scenario by the moderator, we will welcome short interventions (up to a few minutes) by the providers of VLOPs and VLOSEs to illustrate possible approaches to these risks.
We expect a diverse range of CSOs, academics, and providers to participate in the workshop; we can share a list of participants closer to the date of the event, and once we have received confirmations.
The workshop will include a combination of plenary sessions and breakout groups. See below a preliminary agenda (subject to change)

8:30-9:30h: Registration and breakfast
9:30-9:45h: Welcome and opening by European Commission
9:45-10:00h: Presentation on the purpose of the workshop and organisation of sessions
10:00-11:30h: First breakout session
11:30-12:00h: Coffee break
12:00-13:00h: Reporting from the first breakout session and plenary discussion
13:00-14:15h: Lunch
14:15-15:45h: Second breakout session
15:45-16:15h: Coffee break
16:15-17:15h: Reporting from the second breakout session and plenary discussion
17:15-17:30h: Closing plenary and closing remarks by European Commission
17:30-18:30h: Networking cocktail

We aim to publish a short summary of the event shortly after it takes place.
We hope this information helps you plan and prepare for the workshop.
Participants have the possibility to propose aspects for the discussion as part of their registration, and we will take the proposals into account in finalising the agenda. Should you wish to adjust who will participate for your VLOPs and

VLOSEs after the registration period and in light of the more detailed information once available, we are open to accommodate such changes.
Please let us know if you have any further questions or require any additional information.

Best regards,
Organising Team

AMZN_HJC_FCM_0000041

Exhibit 205

DSA MULTI-STAKEHOLDER WORKSHOP ON SYSTEMIC RISKS AND THEIR MITIGATION

7 May 2025

Background information

Risk management is at the core of the Digital Services Act (DSA), obliging providers of very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) to diligently identify, analyse and assess systemic risks, and to use measures for their mitigation that are effective, reasonable and proportionate. Articles 34 and 35 of the DSA cover systemic risks ranging from the dissemination of illegal content and risks to fundamental rights and civic discourse, to risks for mental well-being or risks to children.

The engagement of providers of VLOPs and VLOSEs with stakeholders and experts is important for the good functioning of the DSA. Recital 90 of the DSA encourages providers of VLOPs and VLOSEs to consult representatives of civil society organisations (CSOs), users, or other independent experts when conducting their risk assessments. The published risk assessment reports, as well as other reporting resources pursuant to Article 42(4) DSA, constitute the main means for the public to understand how providers of VLOPs and VLOSEs identify, analyse, assess and mitigate systemic risks stemming from their services.

The DSA multi-stakeholder workshop on systemic risks – by invitation and under Chatham House rule – will bring together the providers of VLOPs and VLOSEs, stakeholders from CSOs and academia, in the presence of the Commission and Digital Services Coordinators (DSCs) to discuss the assessment and mitigation of specific systemic risks. The workshop is an opportunity for the providers of VLOPs and VLOSEs to explain their risk assessment work to stakeholders and to obtain high-quality feedback and ideas.

The multi-stakeholder workshop will be also an opportunity to inform the work on the report on prominent and recurrent systemic risks, and their mitigation by the Board and the Commission as required in Article 35(2) DSA. The first edition of this report will be published later this year and covers the first year of full application of the DSA, spanning the period from 17 February 2024 to 16 February 2025.

Venue

The workshop will take place at the European Commission – CCAB Centre de Conference Albert Borschette, Rue Froissart 36, 1049 Brussels.

Practical Information

- Registration starts at 8:30. Kindly arrive 15 minutes before the event starts, as the security check and onsite accreditation may take some time.
- Upon arrival, please present your ID to the security staff. Afterward, proceed through the security check.
- Breakfast will be served outside the plenary room 0.D. The first plenary session will take place in room 0.D.
- Breakout sessions will take place in room 1.A for Track 1, 3.C for Track 2, 3.A for track 3 and 3.D for Track 4.
- Coffees will be served outside the track rooms 1.A, 3.C, 3.A, and 3.D respectively.
- Lunch and drinks will be served on the 5th floor of the building.
- Reporting of group findings and plenary discussion will take place in the main room 0.D.

Wi-Fi

- **network**: [redacted]
- **login**: [redacted]
- **password**: [redacted]

Agenda

After an introduction in the plenary session, participants will be divided into four tracks, happening in parallel, focusing on different systemic risks categories:

- Track 1 - Dissemination of illegal content
- Track 2 - Civic discourse and elections
- Track 3 - Protection of minors and mental health
- Track 4 - Consumer protection

Each session of the four tracks will be structured around different scenarios that depict different risks along with some questions, which you can find below. These scenarios will guide the participants through a proactive discussion on approaches to identify, analyse, assess and mitigate specific risks.

08.30-09.30	*Registration and breakfast*
09.30-09.45	**Welcome and opening** by [redacted], Deputy Head of Digital Services Unit, DG Connect
09.45-10.00	Presentation of the purpose of the workshop and organization of the sessions
10.00-11.30	First breakout session
11.30-12.00	*Coffee break*
12.00-13.00	Reporting from first breakout session and plenary discussion
13.00-14.15	*Lunch*
14.15-15.45	Second breakout session
15.45-16.15	*Coffee break*
16.15-17.15	Reporting from second breakout session and plenary discussion
17.15-17.30	**Closing remarks** by [redacted], Head of Digital Services Unit, DG Connect
17.30-18.30	*Networking cocktail*

Rules of Engagement

To ensure a productive and respectful discussion, please observe the following engagement rules:

- To ensure that every participant feels comfortable sharing their views and contributing to the discussion in the most meaningful way possible, this meeting will be held under the Chatham House rule.
- The focus of the event is on the risk assessments of providers of VLOPS and VLOSEs, not on enforcement activities of the Commission, which will not be discussed. The Commission will not take any positions on the compliance of regulated entities during the event.
- Please refrain from sharing business-sensitive information during the discussion.
- When participating in the discussion, please state your name and organisation at the beginning of your intervention.
- Please keep your questions and comments brief to allow for a productive discussion.
- Please ensure that your questions and comments are directly related to the DSA risk assessments and risk mitigation measures, and the topic being discussed in the respective track.
- Due to time constraints, we may not be able to take all comments and questions. If you should not have the opportunity to share your thoughts during the workshop, you can send them to [redacted]@ec.europa.eu.

Communicating about the workshop

The workshop has been publicly announced by the European Commission and will be accompanied by public communication outputs, in full respect of the Chatham House Rule referred to above. These may include, but are not limited to, an event summary or similar communication assets, as well as social media posts. When communicating about the event to external audiences,

- Do not: describe the exercise scenarios / name participants / attribute comments to participants without permission
- You can: interview and use quotes from individuals if given explicit permission / talk about the overall topic of the workshop and the tracks / take photos of the event on the condition that persons in the photos agree and no confidential information (such as the scenario) is shown in the photo

List of participating organisations

Track 1 - Dissemination of illegal content	Track 2 - Civic discourse and elections
Accessnow	Amsterdam School of Communication Research (ASCoR)
Apple	AT - RTR
ARTICLE 19	Atlantic Council's Democracy + Tech Initiative
AWO	CERRE
Aylo Freesites Ltd	CheckFirst
Centre for Democracy and Technology Europe (CDT Europe)	Civil Liberties Union for Europe
Conservatoire nationale des Arts et Metiers (Cnam)	DE - BNetzA
Council on Tech and Social Cohesion	EDMO
Coventry University	EE - Consumer Protection and Technical Regulatory Authority
CY - Cyprus RadioTelevision Authority	EFCSN
EL - Hellenic Telecommunications and Post Commission (EETT)	EU DisinfoLab
Electronic Frontier Foundation	European Partnership for Democracy
European Center for Not-for-Profit Law	FI - Traficom
Google	Global Witness
HateAid gGmbH	Google
HU - National Media- and Infocommunications Authority	HR - HAKOM - Croatian Regulatory Authority for Network Industries
IE - Coimisiun na Mean	Institute for Strategic Dialogue (ISD)
Institute for Strategic Dialogue (ISD)	Integrity Institute
Integrity Institute	IT - AGCOM
International Network Against Cyber Hate	Meta
LT - Communications Regulatory Authority of the Republic of Lithuania	Microsoft
Meta	NL - Dutch Authority for Consumers & Markets (ACM)
Microsoft	Oversight Board
MT - Malta Communications Authority	Pagella Politica-Facta

PPMI	PT - ANACOM
Reset Tech	Queens University Belfast
Terre des Hommes Netherlands	RO - National Authority for Management and Regulation in Communications
The Global Disinformation Index	SE - Swedish Post and Telecom Authority (PTS)
University of Amsterdam	SK - Council for Media Services
X	Spark Legal and Policy Consulting
	TikTok
	University of Amsterdam
	University of East Anglia
	Weizenbaum Insitute
	Wikipedia
	X
Track 3 - Protection of minors / mental health	**Track 4 - Consumer protection**
5Rights Foundation	AliExpress
Amnesty International	Amazon
Apple	BEUC
Avaaz	Booking.com
BE - BIPT	CZ - Permanent Representation of the CZ
BEUC	Dansk Erhverv - Danish Chamber of Commerce
BornsVilkar	Erasmus University Rotterdam
Center for Countering Digital Hate (CCDH)	ES - CNMC
Centre for Democracy and Technology Europe (CDT Europe)	FR - Arcom
COFACE	Google
Digital Opportunities Foundation, Germany	LinkedIn
DK - Agency for Digital Government (Digitaliseringsstyrelsen)	Meta
EDRi	Pinterest
Electronic Frontier Foundation	React
Eurochild	Shein
Ghent University	Snap Inc.
Google	Temu
KU Leuven	The Danish Consumer Council
Leiden University	TikTok

LU - Competition Authority	Toy Industries of Europe
Meta	University of Exeter
Microsoft	X
Open Evidence	Zalando
Panoptykon Foundation	
Sciences Po Paris	
Shein	
SI - Agency for Communication Networks and Services of the Republic of Slovenia	
Snap Inc.	
Syddansk Universitets	
TikTok	
University of Amsterdam	
Verbraucherzentrale Bundesverband e.V.	
X	

Scenarios and Questions

TRACK: 1 – DISSEMINATION OF ILLEGAL CONTENT – MORNING SESSION	
Scenario	**5 min**
Luke is a 19-year-old man who has just moved to a new town to study and is trying hard to adjust to the new surroundings. He is struggling to make friends at university, and he often uses the internet as a form of escape, spending hours browsing social media platforms and online forums. One day, Luke comes across a group on platform Delta that seems to share his feelings of frustration and disillusionment. Luke begins to engage with the group, as the group's leaders use rhetoric that resonates with Luke's feelings of anger and frustration, particularly towards women. They begin to lure him in, slowly introducing him to terrorist and extremist ideologies and encouraging him to adopt their views. Soon Luke, thrilled to have finally found a group of men that seems to understand him, starts to share their posts and videos on his social media platforms and even begins to engage in discussion on how to organise offline violence against women. Some of these discussions take place in messaging apps or gaming platforms, but anyone can join using the links shared on platform Delta. In addition, the posts often employ in-group language or slang terms to conceal the violent intentions.	
Risks	**5 min**
Luke faces the risk of being exposed to and disseminating illegal content, particularly terrorism, violent extremism and gender-based violence. Luke is exposed to risks of radicalisation, possibly enhanced by echo chambers and filter bubbles that reinforce his believes by prioritizing content that aligns with his interests and engagements. His use of coded language also increases the possible spread of the harmful content by evading content moderation. Finally, there is a risk of offline violence organised and enabled online.	
Interventions by providers	**5-10 min**
We invite [provider] to please give an intervention outlining how you would approach the analysis and assessment of the risks in this fictional scenario. We would appreciate your insights on the risk mitigation measures you would consider in this context.	
Questions	**60 min**
These questions are intended to support the preparation of participants for the event. During the event, the moderators will facilitate the discussion and invite participants to react to one another's interventions rather than strictly following the questions below. ▪ How could platform Delta analyse and assess whether and how the design, features or functioning of their platform played a role in the identified risks? ▪ How could platform Delta analyse, assess and mitigate the risk of sanctioned entities exploiting the design, features and functioning of platform Delta to disseminate terrorist content? And how can one mitigate the risk of supporters of terrorist groups exploiting the design, features and functioning of platform Delta to disseminate terrorist content?	

- How could platform Delta analyse, assess and mitigate the risks related to outlinks to fringe platforms (e.g. online fora or messaging services)?
- How could platform Delta analyse, assess and mitigate the risk of users exploiting the recommender system of its platform to disseminate and amplify terrorist content?
- How could platform Delta mitigate the risk of echo chambers and filter bubbles?
- What internal processes could platform Delta put in place to (i) support the analysis and assessment of terrorist content-related risks, (ii) support content moderation and the prevention of terrorist content on its platform, (iii) and how can they test and evaluate these processes and their effectiveness?
- What kind of cooperation (e.g. databases, fora) would be beneficial for a swift exchange of information on emerging trends, for example across different EU Member States?
- What awareness-raising measures could platform Delta put in place to educate users like Luke about the risks of extremist content and the importance of reporting suspicious activity?
- What additional measures should be put in place for minors in these situations? How could platform Delta assess and analyse the risks stemming from the interlink of different crime areas on minors?

TRACK: 1 – DISSEMINATION OF ILLEGAL CONTENT – AFTERNOON SESSION	
Scenario	**5 min**
Amira is a 16-year-old Muslim girl who has a history of feeling self-conscious about her identity and has struggled with online harassment in the past. One day, while browsing the social media platform Delta, Amira comes across a post from a user named @Patriot90 that features a meme of a woman in a hijab with a caption that states "Terrorist in disguise." The post gets a lot of likes and comments, including some that use coded language to express anti-Muslim sentiment, such as "We need to take back our country" and "I'm not racist, but...". Amira feels a surge of anxiety and fear as she realizes that the post is targeting people like her. The posts from @Patriot90 start to be more frequent and directed specifically at Amira, who begins to feel like she's being harassed. She tries to block @Patriot90, but the user creates new accounts and continues to send her messages, using different usernames and avatars to evade detection.	
Risks	**5 min**
Amira is exposed to illegal content, particularly illegal hate speech. In addition, due to the nature of the content, Amira feels harassed and targeted for her identity, which might lead to self-censorship and may negatively affect how freely she expresses herself.	
Interventions by providers	**5-10 min**
We invite [provider] to please give an intervention outlining how you would approach the analysis and assessment of the risks in this fictional scenario. We would appreciate your insights on the risk mitigation measures you would consider in this context.	
Questions	**60 min**

These questions are intended to support the preparation of participants for the event. During the event, the moderators will facilitate the discussion and invite participants to react to one another's interventions rather than strictly following the questions below.

- How could platform Delta analyse and assess how the design, features and functioning of their platform influence the dissemination of illegal hate speech?
- How could platform Delta analyse and assess how its terms and conditions influence the dissemination of illegal content on their platform?
- What processes should platform Delta have in place to review and update terms and conditions based on the risks they identified on their platform?
- In the risk analysis and assessment, how can platform Delta consider specific regional or linguistic aspects, for example those specific to Member States?
- How can content moderation processes address the use of coded language or memes that may be used to spread hate speech or discriminatory ideologies?
- How could platform Delta analyse and assess the risk of false positives and false negatives when moderating illegal content (both for automated and human reviewed content)?
- What methods could platform Delta use to evaluate user adoption and engagement with the reporting tools for illegal hate speech, especially among minors?
- How could platform Delta cooperate with trusted flaggers, other providers, or civil society organizations to detect and prevent the spread of illegal content?
- How can in-platform awareness-raising measures be designed and implemented to effectively prevent the spread of illegal hate speech? What methods could platform Delta use to evaluate the effectiveness of tools such as "Kindness Reminders" to curb the dissemination of illegal hate speech?
- How can platform Delta analyse and assess how manipulated imagery, such as deepfakes or AI-generated content, are used to spread hate speech or discriminatory ideologies?

TRACK: 2 – CIVIC DISCOURSE AND ELECTIONS – MORNING SESSION	
Scenario	**5 min**
There are general elections held in an EU Member State and several political candidates face harassment on platform Delta, in the form of private messages to the political candidates and their parties, comments under their posts, as well as content distributed. There are elements of gender-based harassment. CSOs report on suspicions of inauthentic accounts, including potentially automated accounts, being used to promote these messages. As the election day draws closer, the level of harassment is increasing with direct real-life threats against the candidates. Some of the candidates who've been harassed during the campaign, eventually win seats in the election, which exacerbates the harassment and threats, and calls for violence against these candidates start appearing on the platform. Researchers continue to report on potential inauthentic activity in the escalation of harassment.	
Risks	**5 min**
Session focussed on systemic risks in the category of 'any actual or foreseeable negative effects on civic discourse and electoral processes. This is not just mis- or disinformation. This category can also include risks for harassment and even threats for political candidates and office holders. Specific functionalities of platforms can also be abused in such situations, for example through inauthentic accounts or coordinated inauthentic behaviour.	
Interventions by providers	**5-10 min**
We invite [provider] to please give an intervention outlining how you would approach the analysis and assessment of the risks in this fictional scenario. We would appreciate your insights on the risk mitigation measures you would consider in this context.	
Questions	**60 min**
These questions are intended to support the preparation of participants for the event. During the event, the moderators will facilitate the discussion and invite participants to react to one another's interventions rather than strictly following the questions below. ▪ How could platform Delta identify which possible systemic risks could fall under the category 'any actual or foreseeable negative effects on civic discourse and electoral processes? ▪ How could such a process include input from CSOs, academia, political actors or other stakeholders? ▪ How could such a process ensure that the risk assessment is based on the best available information and scientific insights? ▪ How could platform Delta account for geographic/regional context specific variations of the tactics, techniques and procedures malign actors use? ▪ Future proofing risk assessments: how could platform Delta take into account evolving tactics, techniques and procedures in this risk category and how could they anticipate them from posing a risk on their platform?	

- How could platform Delta analyse and assess whether and how the design, features or functioning of their platform played a role in or aggravated the identified risks?
- When designing new (or updating) content moderation policies/guidelines, how could platform Delta take potential risks into account? What guidelines and practices could platform Delta consider when doing so? How could platform Delta ensure in this scenario that their content moderation policies contribute to mitigating risks without infringing upon fundamental rights?
- Narratives (incl. those on harassment) are often comparable across Member States or show specific returning trends (such as calling female politicians witches). What measures could platform Delta put in place to analyse and assess the impact and dissemination of such narratives on their platform to inform their content moderation decisions?
- What tools could platform Delta implement (e.g. reporting centres) to better protect political candidates during campaigning? How could platform Delta assess their effectiveness?

TRACK: 2 – CIVIC DISCOURSE AND ELECTIONS – AFTERNOON SESSION

Scenario	**5 min**

Three weeks before an election, false information about the elections and the electoral process (inaccurate date and location, the elections being cancelled, signing the ballot paper) are being circulated on platform Delta. These messages are amplified by accounts affiliated to influential individuals with a large following on the platform. These individuals are both domestic and foreign. The false claims also get picked up by accounts belonging to foreign state-controlled outlets.

After the elections are over, the same accounts promote messages that dispute the outcome of the elections.

Risks	**5 min**

Risks to votes being invalid or people not expressing their right to vote, and societal risks for covert (foreign) influencing of the election process and undermining the right to free and fair elections. Distrust created by false online accusations may even lead to offline violence.

Interventions by providers	**5-10 min**

We invite [provider] to please give an intervention outlining how you would approach the analysis and assessment of the risks in this fictional scenario. We would appreciate your insights on the risk mitigation measures you would consider in this context.

Questions	**60 min**

These questions are intended to support the preparation of participants for the event. During the event, the moderators will facilitate the discussion and invite participants to react to one another's interventions rather than strictly following the questions below.

- How could platform Delta analyse and assess local risks and context specific risks for a particular election and what does this mean for applicable risk mitigation measures?
- How could platform Delta analyse and assess risks of coordinated inauthentic behaviour used to game the recommender systems in the context of political campaigns?

- How could platform Delta analyse and assess the risk of coordinated inauthentic behaviour for increased content exposure?
- How could platform Delta analyse and assess what are (if any) the risks influential accounts might pose during electoral periods and how could platform Delta decide if and how they have to adapt their platform's recommender systems/content moderation policies to mitigate those risks?
- How could platform Delta analyse and assess the efficacy of additional resources during election periods (e.g. internal "war rooms" or additional measures in content moderation) to boost the capacity to mitigate electoral risks?
- Could platform Delta analyse and assess the impact of promoting authoritative information via in-platform educational offerings during election periods? If so, how could they measure their capacity to mitigate systemic risks to electoral processes?
- How could platform Delta analyse and assess the timing of specific incidents and what could this timing mean for the risk level (e.g. realistic deepfake/audio can have comparatively stronger effects on the success of a candidate on the election date)?
- How could platform Delta analyse assess impersonation of key public figures via manipulated imagery as a systemic risk? If they put mitigation measures in place to address this, how could they evaluate their effectiveness?

TRACK: 3 – PROTECTION OF MINORS AND MENTAL HEALTH – MORNING SESSION	
Scenario	**5 min**
Michael is an introvert 13 years old who struggles with social contacts. He recently changed schools and city and is having a hard time making new friends and connecting with his new classmates. Because of this situation, Michael ends up spending a lot of time alone in his room. He use platform Delta to remain in contact with old friends and classmates and connect with new ones. After connecting with a few new classmates, an account is suggested to him, with a picture of a boy his age who appears to be from the same school and city. Michael starts chatting with his new friend, who also invites him to play online games together and to exchange contacts on other platforms and number-based messaging apps. They start exchanging messages and videochatting via the phone and the platform, until the man behind the username reveals his real name, Jon. Michael knows that he is an older man, but since he seems to be his closest friend, he does not worry about the age difference. After being in contact for months, Jon starts sending sexual and explicit photos and videos to Michael through the platform and other channels, as well as asking Michael to share sexual photos and videos too. Michael starts to feel uncomfortable and scared about the situation, but he is afraid to tell anyone, and does not want to share these feelings with Jon, as he is scared that he will end their contacts.	
Risks	**5 min**
Risks to children's rights (physical integrity, freedom from exploitation and abuse), negative impacts on mental health and wellbeing, as well as proliferation of illegal material and activity (CSAM).	
Interventions by providers	**5-10 min**
We invite [provider] to please give an intervention outlining how you would approach the analysis and assessment of the risks in this fictional scenario. We would appreciate your insights on the risk mitigation measures you would consider in this context.	
Questions	**60 min**

These questions are intended to support the preparation of participants for the event. During the event, the moderators will facilitate the discussion and invite participants to react to one another's interventions rather than strictly following the questions below.

- What features of the service may have enabled Michael to end up in this situation, how might Jon be exploiting these features to make contact, and how would you assess the risk of such features being misused?
- What other conducts may be covered in T&Cs and community rules, besides CSAM, that would be relevant?
- Is content moderation possible in all parts of the platform Delta, and if not, why?
- How can Michael report contacts, behaviour, content that they are uncomfortable with?
- How are contacts on platform Delta potentially being recommended to Michael?
- How can platform Delta analyse and assess functions and features in view of these issues (e.g. contact sync)?
- Should the providers of platform Delta contact any children or organisations representing the interests of children to support them in assessing the risks set out in the fictional scenario?
- Are there any limits for Jon to send manipulated imagery to Michael / Michael's imagery to be safe from download and manipulation?

TRACK: 3 – PROTECTION OF MINORS AND MENTAL HEALTH – AFTERNOON SESSION

Scenario	**5 min**

Joanne is a 17-year-old girl who lives in a small village. She does not have any known mental health issues. She likes travelling, literature, cinema, food and fashion. On platform Delta, Joanne engages with content that her school mates and friends share, and she proactively searches content and follows contacts that she finds amusing or informative according to her interests.

In recent months, she has been worried because some friends of hers seem sad and gloomy. Joanne has also been under pressure at school because of many exams and task.

Joanne looks up what her friends may be facing in terms of mental health struggles. She also looks for ideas for new recipes to cook, places to travel and clothes. She starts being recommended content that makes her wonder about her life's worth, comparing it to the idyllic image of places and people she sees online, as well as content related to diets to look better, anxiety and depression, and even content which suggests self-harm and suicidal ideation as solutions to mental health issues, or challenges related to high-risk self-harm activities. She does not actively engage with or seek the content, other than watching it.

The amount and frequency of this content increases day after day, and Joanne's online activities become longer, increasingly passive and limited to content related to anxiety and depression, as well as challenges inciting harmful behaviours including self-harm. Her mood significantly deteriorates, and due to the increasing time spent on the platforms her school performance worsens substantially.

Joanne does not talk to her parents about her online activities, which does not raise concerns in her family, as this is considered common among adolescents. Consequently, she feels increasingly lonely and drawn to the negative content.

Risks	**5 min**

Risks to children's rights (to safety and development), of negative impacts on mental health and physical wellbeing of minors, and societal risks of minors isolating themselves and not developing or participating in education paths and public life.	
Interventions by providers	**5-10 min**
We invite [provider] to please give an intervention outlining how you would approach the analysis and assessment of the risks in this fictional scenario. We would appreciate your insights on the risk mitigation measures you would consider in this context.	
Questions	**60 min**

These questions are intended to support the preparation of participants for the event. During the event, the moderators will facilitate the discussion and invite participants to react to one another's interventions rather than strictly following the questions below.

- How can platform Delta analyse and assess the risk that minors may make excessive use of social media or display addictive behaviours, to the extent that their development is affected?
- How can platform Delta analyse and assess the risk that minors see accumulations of content that has the potential to negatively affect their development?
- Which features could help Joanne change her experience (if any)? / What features could the provider of platform Delta put in place to help Joanne change her experience?
- How can platform Delta cover any of the issues Joanne is facing by T&Cs or community guidelines?
- How could content be moderated on the platform Delta? Even if the content is not illegal, it can be harmful to minors.
- Are there any additional tools to report content depicting risks presenting specific characteristics (e.g. harmful challenges inciting to self-harm going viral and/or live self-harm content) and/or specifically dedicated to minors?
- How could platform Delta change the algorithmic systems to avoid that the content presented to children does not become increasingly problematic and/or its frequence and accumulation becomes problematic?
- Should the provider of platform Delta cooperate with stakeholders who have experience in mental health issues? If so, how?
- Should platform Delta have any age assurance measures, or by other means inferring it, to tailor recommendations to the user?

TRACK: 4 – CONSUMER PROTECTION – MORNING SESSION	
Scenario	**5 min**
Astrid is 17 and about to finish school. She follows fitness and beauty influencers on a popular platform Delta, and regularly watches their "morning routines" and product reviews. She worries about her weight and sometimes searches for information about dieting and weight loss on other platforms too. In the past, she suffered from an eating disorder. Lately, her feed has been filled with short, aesthetically pleasing videos showing influencers talking about "natural detox" drinks that helped them lose weight fast. The videos are not marked as ads, but they all promote the same link in the bio. She clicked on a couple of the videos out of curiosity, but decided not to buy the product that was available via the link because she wasn't sure it was a good idea. Since then, she sees the videos even more often and sees similar content on other platforms. Eventually, curious and insecure about her body, Astrid clicks the link which leads her to an online marketplace Gamma. Astrid orders the detox product despite its high price, trusting influencers in the videos shown in her feed and positive comments below them. After a week of using the product, she starts having severe stomach pain. She tries to reach the seller on the online marketplace but there are missing contact details of the seller, and she cannot find the way how to contact or notify the online marketplace of this product as she cannot get through automated customer service. The influencer also does not respond. Some of her friends tell her they started seeing the same videos around the time she bought the product, and she is worried they want to try it too. She wants to complain to the platform Delta and stop the bad product being promoted to others, but she doesn't know who is responsible for these links being promoted or where to report it. Astrid feels tricked—she trusted the influencers, not realising the videos were paid for and that they might be inauthentic. She doesn't know who is responsible and where to seek the remedy: the seller? the influencer? one of the platforms? the advertiser?	
Risks	**5 min**
This scenario presents several systemic risks, particularly around digital marketing and content regulation on platforms. First, influencers promote products without disclosing paid partnerships, potentially misleading users like Astrid into buying harmful or ineffective products. Second, there is a lack of transparency and accountability for the promoted product, making it difficult for Astrid to identify who is responsible for adverse effects or misleading claims – whether it's the influencer, the platform, the online marketplace, or the product manufacturer. Third, algorithmic amplification poses a risk, as Astrid keeps seeing more similar content due to her initial interactions, creating a feedback loop that increases her exposure to potentially harmful content.	
Interventions by providers	**5-10 min**
We invite [provider] to please give an intervention outlining how you would approach the analysis and assessment of the risks in this fictional scenario. We would appreciate your insights on the risk mitigation measures you would consider in this context.	
Questions	**60 min**

These questions are intended to support the preparation of participants for the event. During the event, the moderators will facilitate the discussion and invite participants to react to one another's interventions rather than strictly following the questions below.

- What are the main challenges Astrid faces when she realizes the product may be harmful, and how do these challenges originate?
- How can the platforms Delta or Gamma better support users like Astrid in recognizing and avoiding misleading products?
- Would it be different if the product were advertised through the platform's advertising channels?
- How can platform Delta analyse and assess the risk of ad systems directing scam ads to vulnerable consumers based on data which indicates that they are more likely to take action/click on this type of content (e.g. browsing history, previous engagement with other types of scam content etc.)?
- How can the platforms Delta or Gamma ensure their Terms & Conditions effectively address issues of undisclosed advertisements or misleading product claims?
- What can the platforms Delta and Gamma do to ensure the enforcement of these Terms & Conditions to protect consumers like Astrid?
- How can platform Delta differentiate between genuine user-generated content and undisclosed advertisements?
- How can platform Delta effectively collaborate with trusted flaggers to mitigate risks related to undisclosed or misleading marketing?
- Could cooperation between online marketplaces and other online platforms play a role in addressing these issues?
- How can platform Delta protect users from misleading content generated by means of AI systems, particularly in the context of product marketing?

TRACK: 4 – CONSUMER PROTECTION – AFTERNOON SESSION

Scenario	**5 min**

Elisa is a 42-year-old single mother living in a rural area. She relies on online shopping for most of her household needs. When her child develops a mild skin condition, she searches on the popular platform Gamma, for a gentle, natural cream. The top search results look convincing—product descriptions are well-written, and some even mention dermatological studies.

What Elisa doesn't realise is that many of these product pages were automatically generated using AI tools by third-party sellers. The descriptions include fabricated claims, scientific-sounding language, and even fake user Q&As generated to build trust. The platform's moderation tools don't flag the content, because it doesn't trigger typical spam or prohibited-claim filters. The third-party seller used AI to generate the content because he did not have required detailed information about that product, including a safety report on the composition of the product and its instructions for use as it was not produced in line with the EU legislation. After using the cream, Elisa's child has an allergic reaction. She discovers that the product's ingredients aren't clearly listed, and that the seller is using a shell company address. She also realises that some reviews were AI-written and copy-pasted across multiple products.

Risks	**5 min**

Elisa's child is facing health related issues due to the unsafe product bought online by their mother who relies on online shopping. The main risks in this scenario are risks to consumer protection, protection of public health and negative consequences to the person's physical well-being. Elisa is also at the point of losing her trust in online shopping which could complicate her life as physical shops are more difficult to reach for her.

Interventions by providers	**5-10 min**

We invite [provider] to please give an intervention outlining how you would approach the analysis and assessment of the risks in this fictional scenario. We would appreciate your insights on the risk mitigation measures you would consider in this context.

Questions	**60 min**

These questions are intended to support the preparation of participants for the event. During the event, the moderators will facilitate the discussion and invite participants to react to one another's interventions rather than strictly following the questions below.

- How can platform Gamma analyse and assess how the design, features and functioning of their platform influence the dissemination of illegal content, such as illegal/unsafe goods or unauthentic product descriptions?
- How can platform Gamma make sure that its applicable Terms & Conditions are effectively enforced vis-a-vis users' generated content such as false reviews or comments which might mislead customers like Elisa into buying illegal or dangerous products?
- How can platform Gamma assess and improve the functioning of its recommender systems in a way that trustworthy and safe products would be given priority?
- How can platform Gamma analyse and assess risks of disseminating illegal/harmful products when bringing new features to the platform, such as AI assistants?
- How can platform Gamma facilitate the knowledge and expertise of trusted flaggers and/or CSOs (e.g. consumer protection organisations) into its risk assessment and mitigation measures?
- How can platform Gamma establish/maintain/deepen their relationships with trusted flaggers and/or consumer protection organisations to improve the detection and prevention of the most prominent risks on online marketplaces, such as e.g. the dissemination of illegal content?

Exhibit 206

European Commission - DSA Systemic Risk Assessment Workshop Readout

May 7, 2025

- On May 7th, the European Commission (EC) hosted a DSA Systemic Risk Assessment (SRA) Workshop that included: Civil Society Organisations (CSOs), Digital Services Coordinators (DSCs) and some Very Large Online Platforms (VLOPs) and Very Large Search Engines (VLOSEs). The VLOPs were: AliExpress, Apple, Amazon, Booking.com, Google, LinkedIn, Meta, Microsoft, Pinterest, Pornhub, Shein, Snap, Temu, TikTok, X, and Zalando. The event was organized around four simultaneous breakout sessions covering: dissemination of illegal content, civic discourse and elections, protection of minors, and consumer protection.

- Initial remarks were provided by Marco Giorello, Head of Unit, DG Connect regarding industry SRA reports. Giorello noted, in part:

    - The EC views EU civil society not only as a source of substantive knowledge but also as an important player in assessing the success of the DSA's implementation and its broader enforcement. By referencing design issues and data to support platform risk mitigations, which are focus points for many NGOs, the Commission gave evidence of listening closely to civil society on matters of systemic risk. That said, the Commission explicitly conceded that it was not realistic to expect platforms to conduct 1:1 engagements with multiple civil society groups. The Commission offered no substantive suggestion on how to conduct CSO engagement.

    - There is no plan to issue guidelines or formal feedback on Articles 34 and 35; EC feedback should be deduced from RFIs which indicate where more information is expected. The EC stated the intention for SRAs to be compared year-over-year.

    - Pursuant to Art 35(2), the European Board for Digital Services and the Commission are required to publish a report every year of the most prominent DSA risks. This year's report will cover the period from 17 Feb 2024 - 16 Feb 2025 and will cover European Union and national level risks and the mitigation measures; the publication date was not specified. Civil Society have contributed to this project, their feedback is being analysed.

    - SRAs should be focused on transparency and providers should not only publish results of assessments but also set out how they reached the results. Providers should also include the relevant results of all the risks and the risk factors (not only a subset of those).

    - The European Commission has noted CSOs' feedback that Reports focus on the content on the platforms but do not sufficiently cover the risks related to platform design, which is a requirement for SRAs pursuant to Article 34(2). The EC did acknowledge that the Reports are of necessity evolving, given the DSA's nasency.

PRODUCED TO HJC

- SRAs need to set out the evidence behind statements made; for example, where policies are mentioned as mitigation measures, platforms should outline the result of the deployment of these measures.
- SRAs should explain how a service is designed and functions so the public can understand the risk assessment.
- Given the current stage of technology and the reliance on AI by all platforms, SRAs should have specific assessments concerning Generative AI risks - “It is clear and very important that the RAs contain a beefed up and specific assessment of the risks identified by AI generated content”.
- As required by Article 34, the assessments must also analyse whether and how the risks are influenced by intentional manipulation of their service.

- **Dissemination of Illegal Content**: Initial discussion revolved around data (sought from platforms in relation to their risk mitigations) and design issues. Much of the discussion also merged into dialogue on familiar content moderation issues – such as linguistic challenges, adequacy of reporting channels, and protection for minority groups. There was no strong push by civil society participants for new stakeholder engagement mechanisms to address issues of systemic risk.
  - A summary of EC-provided takeaways from the two sessions:
    - Illegal content is a symptom, a broader risk approach is needed
    - Need for a holistic approach to assessing platform design, policies, etc
    - Best practices: Labelling of harmful content; blacklisted URLs, continuous review of community guidelines, banned keyword lists, coded language, kindness reminders, recommender systems not only based on engagement, labelling
    - Role and importance of industry-wide initiative and information sharing (e.g., GIFCT)
    - Call for enhanced transparency on underlying analysis and evidence
    - Role of reporting tools and trusted flaggers, role can be adapted to risk identified
    - Content moderation needs to be adapted to national legislation and linguistic/regional specificities
- **Civil Discourse and Elections:** A key theme in the discussion was CSOs’ call for access to data and disclosure of underlying evidence behind our risk assessments in order to verify the claims companies make.
  - A summary of EC-provided takeaways from the two sessions:

PRODUCED TO HJC

        - Some CSOs noted that there was a discrepancy between the risk assessments that suggest that mitigations are appropriate and risks levels are low and external studies on online harms that suggest the opposite;
        - It's important to know the evidence base of the assessments, particularly measuring the effectiveness of mitigation measures
        - There is a need for the assessment of language/local context expertise
        - Call for effective data access and transparency, particularly for all Member States
        - The need for cross-platform information sharing (e.g. around CIB, similar to GIFCT) and better communication about risk assessments and mitigations
        - Media pluralism was noted as an area that was underdeveloped or unaddressed in the SRAs
        - Recommender systems were suggested as a mitigation measure to surface authoritative content

- **Protection of Minors**: These sessions were primarily focused on feedback from CSOs, which ranged from outlining observations regarding challenges with having minors on social media, to recommendations for further mitigations.
    - A summary of EC-provided takeaways from the two sessions:
        - Early involvement of experts, CSOs, and children into the risk assessment process
        - Risk assessment focuses should be broader, not only risks but also benefits of new product (CIRA)
        - Improvement areas include more information on effectiveness measures, published data as evidence, and access by researchers. A specific emphasis on the importance of platforms to have reliable age verification
        - Risks assessments should outline benefits to children too; balancing positive outcomes with the risks
        - Mitigations should ensure that safety is implemented considering children's agency, experience, and diversity, and be tailored to the changing risks and platform uses
        - More transparency / data on analysis / assessment and mitigation effectiveness needed for research communities

PRODUCED TO HJC

- **Consumer Protection**: These sessions were attended by a limited number of CSOs and DSCs, but were well-represented by platforms, including all major marketplaces. The discussions did not generate much debate, instead focusing on compliance-driven interventions, particularly from the marketplaces.
    - A summary of EC-provided takeaways from the two sessions:
        - The risk assessment methodologies were similar across platforms; there is a recommendation to involve CSOs, trusted flaggers, and metrics more into the methodology
        - There is a focus on legal obligations
        - The ecosystem is complex across platforms, sellers, and users adapting to new risks from new product types
        - Monetisation and off-platform trends are risk factors (linking e-commerce and social media)
        - There should be a shared responsibility risk management between the seller and platform
        - There is the role of Generative AI as both as a risk factor and a risk mitigation measure

PRODUCED TO HJC

Exhibit 207

EUROPEAN COMMISSION

PROTECTION OF YOUR PERSONAL DATA

This privacy statement provides information about the processing and the protection of your personal data.

Processing operation: DSA Multi-stakeholder workshop on Systemic Risks

Data Controller: European Commission, *DG CNECT, F2*

Record reference: [redacted] on events and meetings

Table of Contents

1. **Introduction**
2. **Why and how do we process your personal data?**
3. **On what legal ground(s) do we process your personal data?**
4. **Which personal data do we collect and further process?**
5. **How long do we keep your personal data?**
6. **How do we protect and safeguard your personal data?**
7. **Who has access to your personal data and to whom is it disclosed?**
8. **What are your rights and how can you exercise them?**
9. **Contact information**
10. **Where to find more detailed information?**

1. Introduction

The European Commission is committed to protect your personal data and to respect your privacy. The Commission collects and further processes personal data pursuant to Regulation (EU) 2018/1725 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data.

This privacy statement explains the reason for the processing of your personal data in the context of **DSA Multi-stakeholder workshop on Systemic Risks**. It explains the way we collect, handle and ensure protection of all personal data provided, how that information is used and what rights you have in relation to your personal data. It also specifies the contact details of the responsible Data Controller with whom you may exercise your rights, the Data Protection Officer and the European Data Protection Supervisor.

The information in relation to processing operation '**DSA Multi-stakeholder workshop on Systemic Risks'** undertaken by **DG CNECT, Unit F2 Digital Services** of the European Commission, is presented below.

2. Why and how do we process your personal data?

Purpose of the processing operation:

DG CNECT, Unit F2 Digital Services collects and further processes your personal data to provide you with information about the DSA Multi-stakeholder workshop on Systemic Risks (before, during and after), to process your registration in those events and to follow up after the event.

Your personal data will not be used for any automated decision-making including profiling.

3. On what legal ground(s) do we process your personal data?

The processing operations linked to the organisation, management, promotion and follow-up of the **DSA Multi-stakeholder workshop on Systemic Risks** are necessary for the management and functioning of the Commission, as mandated by the Treaties. Those provisions are Article 11 of the Treaty on European Union and Article 15 of the Treaty on the Functioning of the European Union. Consequently, those processing operations are lawful under Article 5(1)(a) of Regulation (EU) 2018/1725 (processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the Union institution or body).

4. Which personal data do we collect and further process?

The following personal data will be processed:

- contact details (function/title, first name, last name, name of organisation, e-mail address,).
- nationality, passport or identity card number and its date of issue and expiry date may be collected, so that the data subjects may obtain access to the premises where the meeting/event is held.

5. How long do we keep your personal data?

The Data Controller only keeps your personal data for the time necessary to fulfil the purpose of collection or further processing.

For each of the categories of personal data that is processed, please find below the retention details and the reference to the relevant record of processing:

- All personal data related to the organisation and management of the **DSA Multi-stakeholder workshop on Systemic Risks** will be deleted **five years** at the latest after the last action in relation to the roundtable discussions.
- Personal data shared with the Directorate-General for Human Resources and Security of the European Commission for the participants to gain access to Commission buildings is kept **for 6 months** after the termination of the link between the data subject and the Commission. More information is available in the Record of Processing DPR-EC-00655 (Commission Physical Access Control System (PACS)).

6. How do we protect and safeguard your personal data?

All personal data in electronic format (e-mails, documents, databases, uploaded batches of data, etc.) are stored either on the servers of the European Commission or of its contractors. All processing operations are carried out pursuant to Commission Decision (EU, Euratom) 2017/46 of 10 January 2017 on the security of communication and information systems in the European Commission.

In order to protect your personal data, the Commission has put in place a number of technical and organisational measures. Technical measures include appropriate actions to address online security, risk of data loss, alteration of data or unauthorised access, taking into consideration the risk presented by the processing and the nature of the personal data being processed. Organisational measures include restricting access to the personal data solely to authorised persons with a legitimate need to know for the purposes of this processing operation.

7. Who has access to your personal data and to whom is it disclosed?

Access to your personal data is provided to the Commission staff in DG CNECT Directorate F Unit F2 responsible for carrying out this processing operation and to other authorised Commission staff according to the "need to know" principle. Such staff abide by statutory, and when required, additional confidentiality agreements.

Cookies

Cookies are short text files stored on a user's device (such as a computer, tablet or phone) by a website. Cookies are used for the technical functioning of a website (functional cookies) or for gathering statistics (analytical cookies).

The registration for the **DSA Multi-stakeholder workshop on Systemic Risks** takes place via Event-Works[1]. The cookies employed by the Commission on the registrant's device for that

[1] For more information on the processing of personal data via Event-Works, see DPR-EC-00297 "Participants registration for Commission conferences and events using Event-Works"

purpose will be covered by the cookie policy of the Commission, which is available here: https://ec.europa.eu/info/cookies_en.

Cookies are stored by Europa Analytics, the corporate service which measures the effectiveness and efficiency of the European Commission's websites on EUROPA. More information is available in the Record of Processing DPR-EC-00685 (Europa Analytics).

Enabling these cookies is not strictly necessary for the website to work but it will provide you with a better browsing experience. You can delete or block these cookies, but if you do that, some features of the meeting/event website may not work as intended.

The cookie-related information is not used to identify data subjects personally and the pattern data is fully under the Commission's control. These cookies are not used for any purpose other than those described here.

Should you wish to opt your personal data out of our anonymised, aggregated statistics, you can do so on our cookies page. In particular, you can control and/or delete those cookies as you wish.

International transfers

Please note that pursuant to Article 3(13) of Regulation (EU) 2018/1725 public authorities (e.g. Court of Auditors, EU Court of Justice) which may receive personal data in the framework of a particular inquiry in accordance with Union or Member State law shall not be regarded as recipients. The further processing of those data by those public authorities shall be in compliance with the applicable data protection rules according to the purposes of the processing.

The information we collect will not be given to any third party, except to the extent and for the purpose we may be required to do so by law.

8. What are your rights and how can you exercise them?

You have specific rights as a 'data subject' under Chapter III (Articles 14-25) of Regulation (EU) 2018/1725, in particular the right to access, your personal data and to rectify them in case your personal data are inaccurate or incomplete. Where applicable, you have the right to erase your personal data, to restrict the processing of your personal data, to object to the processing, and the right to data portability.

You have the right to object to the processing of your personal data, which is lawfully carried out pursuant to Article 5(1)(a) of Regulation (EU) 2018/1725 on grounds relating to your particular situation.

In case of conflict, you can contact the Data Protection Officer. If necessary, you can also address the European Data Protection Supervisor. Their contact information is given under Heading 9 below.

Where you wish to exercise your rights in the context of one or several specific processing operations, please provide their description (i.e. their Record reference(s) as specified under Heading 10 below) in your request.

9. Contact information

- **The Data Controller:** If you would like to exercise your rights under Regulation (EU) 2018/1725, or if you have comments, questions or concerns, or if you would like to submit a complaint regarding the collection and use of your personal data, please feel free to contact [redacted]@ec.europa.eu.

- **The Data Protection Officer (DPO) of the Commission:** You may contact the Data Protection Officer ([redacted]@ec.europa.eu) with regard to issues related to the processing of your personal data under Regulation (EU) 2018/1725.

- **The European Data Protection Supervisor (EDPS):** You have the right to have recourse (i.e. you can lodge a complaint) to the European Data Protection Supervisor ([redacted]@edps.europa.eu) if you consider that your rights under Regulation (EU) 2018/1725 have been infringed as a result of the processing of your personal data by the data controller.

10. Where to find more detailed information?

The Commission Data Protection Officer (DPO) publishes the register of all processing operations on personal data by the European Commission, which have been documented and notified to him. You may access the register via the following link: http://ec.europa.eu/dpo-register.

Exhibit 208

From: [redacted]
To:
Subject: Fwd: Invitation to the DSA Multi-stakeholder workshop on Systemic Risks – 7 May 2025
Date: Wednesday, July 2, 2025 4:50:50 PM

---------- Forwarded message ---------
From: [redacted] <[redacted]>
Date: Wed, Jul 2, 2025 at 1:32 PM
Subject: Fwd: Invitation to the DSA Multi-stakeholder workshop on Systemic Risks – 7 May 2025
To: [redacted] <[redacted]>

This thread is the most informative

---------- Forwarded message ---------
From: [redacted] **via** [redacted] <[redacted]>
Date: Fri, 25 Apr 2025 at 16:28
Subject: RE: Invitation to the DSA Multi-stakeholder workshop on Systemic Risks – 7 May 2025
To: [redacted] <[redacted]>
Cc: [redacted] <[redacted]@ec.europa.eu>, [redacted] <[redacted]>, [redacted]
[redacted]@ec.europa.eu <[redacted]@ec.europa.eu>

Dear [redacted],

You should have received additional information last week concerning the structure of the day, which we are attaching below. The four tracks will run in parallel, which means that in the morning / afternoon breakout sessions we will have four rooms devoted to the four tracks, followed by plenary discussions in which participants will reconvene and discuss the insights from each tracks. Early next week we will share the scenarios and questions along with other background information.

Kind regards,

[redacted] for the team

Dear Sir/Madam,

In view of the upcoming DSA Multi-stakeholder workshop on Systemic Risks which will take place on 7 May 2025, we want to give you some additional details on the format of the workshop. The workshop will take place at the European Commission – CCAB Centre de Conference Albert Borschette, Rue Froissart 36, 1049 Brussels.

After an introduction in the plenary session, participants will be divided into four tracks, happening in parallel, focusing on different systemic risks categories: Track 1 - Dissemination of illegal content, Track 2 - Disinformation and elections, Track 3 - Protection of minors and mental health, and Track 4 - Consumer protection.

Overall, the objectives of the workshop are to foster collaboration on risk assessments, help participants learn about the providers' different approaches to the identification, assessment, and mitigation of systemic risks, and gather perspectives from Digital Services Coordinators (DSCs), civil society organisations (CSO), researchers, and providers of VLOPs and VLOSEs.

Each session of the four tracks will be structured around different scenarios that depict different risks. These scenarios will guide the participants through a proactive discussion on approaches to assess and mitigate certain risks. The scenarios as well as the guiding questions will be shared in advance for the participants to be prepared. After an introduction into the specific risk scenario by the moderator, we will welcome short interventions (up to a few minutes) by the providers of VLOPs and VLOSEs to illustrate possible approaches to these risks.

We expect a diverse range of CSOs, academics, and providers to participate in the workshop; we can share a list of participants closer to the date of the event, and once we have received confirmations.

The workshop will include a combination of plenary sessions and breakout groups. See below a preliminary agenda (subject to change)

8:30-9:30h: Registration and breakfast

9:30-9:45h: Welcome and opening by European Commission

9:45-10:00h: Presentation on the purpose of the workshop and organisation of sessions

10:00-11:30h: First breakout session

11:30-12:00h: Coffee break

12:00-13:00h: Reporting from the first breakout session and plenary discussion

13:00-14:15h: Lunch

14:15-15:45h: Second breakout session

15:45-16:15h: Coffee break

16:15-17:15h: Reporting from the second breakout session and plenary discussion

17:15-17:30h: Closing plenary and closing remarks by European Commission

17:30-18:30h: Networking cocktail

We aim to publish a short summary of the event shortly after it takes place.

We hope this information helps you plan and prepare for the workshop.

Participants have the possibility to propose aspects for the discussion as part of their registration, and we will take the proposals into account in finalising the agenda. Should you wish to adjust who will participate for your VLOPs and VLOSEs after the registration period and in light of the more detailed information once available, we are open to accommodate such changes.

Please let us know if you have any further questions or require any additional information.

Best regards,

Organising Team

From: [redacted] <[redacted]>
Sent: Friday, April 25, 2025 3:01 PM
To: [redacted] <[redacted]@ec.europa.eu>
Cc: [redacted] (CNECT) <[redacted]@ec.europa.eu>; [redacted] (CNECT) <[redacted]@ec.europa.eu>; [redacted] <[redacted]>
Subject: Re: Invitation to the DSA Multi-stakeholder workshop on Systemic Risks – 7 May 2025

Dear CNECT team

We are planning on sending a number of team members to the 7 May workshops but are wondering whether they are concurrent or consecutive? Do you have agendas available yet?

Many Thanks

[redacted]

On Fri, Mar 21, 2025 at 8:53 AM [redacted]@ec.europa.eu <[redacted]@ec.europa.eu> wrote:

> Dear Sir, Dear Madam,
>
> We are delighted to invite you to attend in person a workshop that we will hold in Brussels on **7 May 2025,** 9:00-18:30 CET to discuss specific aspects of risk assessments under the Digital Services

Act (DSA).

The workshop, to be held in person and upon invitation, brings together the providers of Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs), stakeholders from civil society organisations (CSO) and academia, and Digital Services Coordinators (DSCs) to discuss the assessment and mitigation of specific systemic risks.

The European Board for Digital Services and the Commission are working on a report on recurrent and prominent systemic risks referred to in Article 35(2) of the DSA, which will cover the first year of full application of the DSA, spanning the period from 17 February 2024 to 16 February 2025. One objective of the workshop is to ensure that this work is informed by the providers' current practices regarding risk assessment and risk mitigation, and to benefit from an open exchange with the Commission, national DSCs, CSO and academics.

In addition, the workshop will take place at a time when many providers of VLOPs and VLOSEs are working on their yearly risk assessments. It is an opportunity for the providers of VLOPs and VLOSEs to obtain high-quality feedback and input, which may be leveraged for ongoing and future risk assessments.

The workshop will be structured around break-out sessions on different systemic risks categories, such as dissemination of illegal content, disinformation and elections, protection of minors and mental health, and consumer protection.

We look forward to your participation. You can confirm the participation of your representative(s) by **registering via this form**. We would be grateful if you confirmed the participation of your representative(s) by 16 April.

Sincerely,

[redacted]

Head of Unit

--
[redacted]

You received this message because you are subscribed to the [redacted]
To unsubscribe from this group and stop receiving emails from it, send an email to [redacted]
To view this discussion visit

Exhibit 209

From: [redacted]
To: [redacted]
Subject: Fwd: DSA Risk Assessment Roundtable - readout
Date: Wednesday, July 2, 2025 4:52:21 PM

---------- Forwarded message ---------
From: [redacted] <[redacted]>
Date: Wed, Jul 2, 2025 at 1:02 PM
Subject: DSA Risk Assessment Roundtable - readout
To: [redacted] <[redacted]>

As for the CSOs that were present at the meeting, I checked and I can't find any list, but I can point at some I remember meeting there:
- ISD Institute for Strategic Dialogue - panel on disinformation, quite aggressive and critical against platforms not working with fact checkers.
- Representative of EDMO (network of EU fact checkers and researchers): the most aggressive (see in the readout)
- Access Now: claiming platforms' content moderation efforts should go beyond illegal content and lead to removal of everything that can be considered as hateful and harmful.
- INACH network: several EU-based network of hate speech organisations. Problematic one is the Never Again Association from Poland. The others are pretty much ok.

The panel on disinformation was definitely the most difficult, because of the presence of fact checkers.

Readout:

Context: On 7 May 2025, the European Commission hosted a significant event in Brussels, convening approximately 200 representatives from Very Large Online Platforms and Search Engines (VLOPSES), Digital Service Coordinators (DSCs), civil society organizations (CSOs), and academia to address the assessment and mitigation of systemic risks. The event was structured into four thematic tracks, focusing on critical issues: the dissemination of illegal content, civic discourse and elections, protection of minors and mental health, and consumer protection. [redacted] participated in all four tracks, contributing to discussions on these pressing challenges.

Key takeaways:

- **Track 1 - Dissemination of illegal content**

In general, there was a tendency from civil society to point at enforcement issues rather than focusing on guidance concerning risk assessments. The EC tried several times to divert the attention from enforcement back to the event's objectives.

Areas addressed: DSA reporting systems, mitigation measures related to recommender systems, issues with evaluation of terrorist content and illegal hate speech.

- CSOs pointed at difficulties in using DSA reporting mechanisms and the fact that it is even more problematic for regular/non expert users.
- CSOs called for more efforts to reduce the spread of TCO and illegal hate speech through recommender systems.
- CSOs claimed moderation efforts must go beyond illegality and also better address harmful content and disinformation aimed at dehumanising or inciting hate.
- Some suggested labelling is not enough when it comes to hate, even if not illegal forms of hate.
- Reflections around how platforms prevent risk of over/under-removal, false positives and false negatives.

- **Track 2 - Civic discourse and elections**
  - **Accessing the API is a cumbersome process for researchers; it typically takes over eight weeks to receive a response from us, yet researchers are then given only two weeks to reply. The process could take months without resolution (This was [redacted]).**
  - **There is limited evidence/data on the proportionality and effectiveness of the mitigation measures and CSOs asked for more information or at least results of tests.**
  - **[redacted]was questioned about the number of moderators and language coverage in relation to addressing election-related risks. Representatives of the EU Digital Media Observatory (EDMO), a network of fact checkers and researchers across the EU, was the most critical during the session, who made claims that CNs do not work. They also pushed quite aggressively the allegation [redacted] lacks necessary resources (content moderation, T&S etc.) to address systemic risks.**

- **What data underpins the platform's risk assessment? There is interest in incorporating external research into the RA process and enabling CSOs to provide feedback.**
- **Civil society is seeking more information on ad-hoc risk assessments and the overall lifecycle of RAs.**
- **Recommendation systems were mentioned as a potential mitigation measure to help surface authoritative content.**
-

- **Track 3 - Protection of minors and mental health**
  - The DSA reporting forms are difficult for both users and researchers to navigate. It takes too long time for users/parents etc. to report (CSO on child safety).
  - High risk functionalities
    - Multiple accounts under the same profile - risk that perpetrators have a fake child profile alongside a regular adult profile.
    - Ability to log-in on the same device to multiple accounts.
    - Fake email addresses - most people have email addresses connected to their name, but bad actor accounts tend to have incoherent email addresses. Can a tool be developed to detect this?
    - Engagement incentives that lead to excessive use.
    - Rabbit holes: independent from harmful content; harm comes from time-spent and cumulative exposure and lack of pluralism.

- Preexisting vulnerabilities and different ages of child users impact the varying effectiveness of time-based controls

- **Track 4 - Consumer Protection:** The first session examined a case involving undisclosed paid partnerships and the promotion of dangerous weight loss drinks, with a focus on [redacted] and [redacted] emphasized its robust mitigation efforts, highlighting its clear Paid Partnership Policy requiring influencers to use "#ad" for transparency and a strict policy prohibiting the promotion of such products. However, CISOs and DSCs pressed for more details on how platforms measure policy effectiveness. The second session explored AI-generated product descriptions on marketplaces, spotlighting [redacted]. CISOs and DSCs expressed concerns about persistent online shopping risks despite regulations, questioning platform safety measures. Platforms acknowledged the challenge of unpredictable bad actor behaviors, noting their ongoing efforts to mitigate risks while clarifying that completely eliminating them is unrealistic.

[redacted]

Exhibit 210

From: [redacted]
To:
Subject: Fwd: Thank you for attending the DSA Multi-stakeholder Workshop on Systemic Risks
Date: Wednesday, July 2, 2025 4:54:24 PM

---------- Forwarded message ---------
From: [redacted] <[redacted]>
Date: Wed, Jul 2, 2025 at 1:40 PM
Subject: Fwd: Thank you for attending the DSA Multi-stakeholder Workshop on Systemic Risks
To: [redacted] <[redacted]>

Here is the high lev readout (not as detailed as ours).
Still looking for the list of participants

---------- Forwarded message ---------
From: [redacted]**@ec.europa.eu** <[redacted]@ec.europa.eu>
Date: Thu, 15 May 2025 at 10:44
Subject: Thank you for attending the DSA Multi-stakeholder Workshop on Systemic Risks
To: [redacted]@ec.europa.eu <[redacted]@ec.europa.eu>

Dear Participant,

We wanted to take a moment to express our gratitude for your participation in the DSA Multi-stakeholder Workshop on Systemic Risks, which took place on 7 May 2025.

Your valuable input and feedback are greatly appreciated. If you have any additional comments or suggestions, please do not hesitate to share them with us at [redacted]@ec.europa.eu.

Ahigh-level summary of the event is now available on the Commission's website: https://digital-strategy.ec.europa.eu/en/news/commission-holds-workshop-platforms-and-civil-society-assessment-online-risks. We hope that the discussions and exchanges that took place during the sessions will contribute to collaborations between stakeholders, in line with the spirit of Recital 90 DSA.

Once again, thank you for your participation and contribution to the workshop.

Best Regards,

The Organising Team

--

[redacted]

Section VII:
Elections Exhibits

Exhibit 211

Slovak Election Summary Report 2023

Over a four month period, TikTok wide cross functional taskforce worked together to prepare for the recent election in Slovakia held on 30 September.

The European Commission (EC) made it clear that social media platforms will be tested on their COPD and DSA commitments during the Slovak election. Furthermore, online platforms were summoned to Bratislava on 12 September for an in-person meeting with the Commission to discuss election risks and preparedness. Also, TikTok participated in COPD elections working group dedicated cross-platform, EC and civil society organisation meetings held to discuss risks, developments and actions taken by platforms.

Timeline

* 30 January 2023: Slovak Parliament announced a general election scheduled for 30 September 2023. As part of TikTok Election Integrity Programme, Slovak election is added to the TikTok's internal Elections Calendar and initial planning work commences as part of the global programme.
* 12 June: Cross-functional (XFN) Slovak Election Working Group convened for detailed planning purposes.
* 30 June: Meeting between Slovak Council for Media Services and members of TikTok's Slovak Election Working Group.
* July-August-September: Ongoing intensive election preparations, scenario-planning and engagement, policy and product work.
* 30 September: Scheduled election day for Slovak general (parliamentary) election.
* 2 October: Election results announced

Partnerships

* We have considered the local risks identified by key stakeholders and experts including the Slovakian Council for Media Services, the Slovak Interior Ministry's Centre for Countering Hybrid Threats and DigiQ Slovak civil society organisation focused on online safety. We have actively cooperated with them on other elements, including on the development of our in-app interventions and review of concerning content via Community Partner Channel.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

* In partnership with DigiQ we developed our media literacy strategy with the objective of educating users on how to detect misinformation and verify information they are seeing online. DigiQ s educational videos were available from our in-app Election Hub and also redirected users to authoritative information on their website. Videos were created on DigiQ TikTok account and will remain available to Slovak users.
* Fact-check partner Lead Stories supported our election integrity efforts in Slovakia by proactively flagging potentially violating content on TikTok as well as by reviewing escalated videos through our dedicated Slovak fact-checking queues.

Product

* The Slovak election voter information center (Elections Hub) was launched on 4 September in collaboration with The Slovak Centre of Media Services. Users had access to essential voter information and media literacy videos produced by our partner DigiQ. Elections hub was accessible from special labels (notice tags) on elections-related videos and from search banners (Search Guide) which appeared when users were searching for elections-related videos. Please find graphical examples in the Appendix of this document.
    - Election Hub visits: 67,390 views by 46,083 unique users
    - Election-related videos labelled: 7371

Election Hub Stats	
Visits to Election Hub (all features)	67,390
Unique user visits to Election Hub (all features)	46,083
Users on average stayed at Election Hub (seconds)	13,8
Media Literacy and Official sources of Information	
Button 1 minv.sk clicks	2,990
Button 2 minv.sk clicks	257
Button 3 DigiQ.sk clicks	90
Text Hyperlink je zapisany clicks	56
Text Hyperlink hlasovacieho preukazu clicks	66
Video 1 play count	188
Video 2 play count	79

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_019054

Video 3 play count	23
Video 4 play count	35
Video Notice tags	
Number of election-related hashtags used for tagging	85
Number of videos with election notice tag for the time period from 1st June to 20th October	7,371

Proactive work

Mitigating risk of deepfake

* Based on CMS deepfake escalations, the T&S team performed continuous media monitoring which resulted in early identification of deepfakes on peer platforms. This enabled us to catch a range of deepfakes before they had a chance to proliferate on TikTok. Deepfakes were published on the platform gradually on Friday, one day prior to the election. Content was targeting Šimečka, leader of Progressives and Čaputová, Slovak president.
* While these deepfakes were widely reported in media a few days after the election, by Bloomberg (paywalled, copy here), another one here or Wired. Thanks to swift actions taken on our side, deepfakes did not get a chance to engulf Slovak TikTok, unlike in the cases of Meta and Youtube.

Proactive elections content monitoring and sampling

Proactive review was conducted based on a specifically designed keywords list. Between 14/09 and 30/09 we proactively reviewed 1870 pieces of content (highest VVs, highest VV growth & videos reported for Election Misinformation, with the action rate of 6%. This speaks both to great moderation quality, as well as effective removal of proactively surfaced violating content.

Escalations

Throughout September we have received and reviewed four external escalations from our partners. Please find the comprehensive rundown below or refer to the escalations tracker.

Escalation 1

* During our meeting in Bratislava, [redacted] mentioned 40 accounts that they believe are in the breach of our CGs.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

* In total, they sent 63 accounts, featuring two duplicates = 61 accounts in total. Of these, 10 were already banned by the time we received this report. 8 of which were banned prior to the meeting in Bratislava.
* Detailed review Slovak elections further accounts review

Review results:

* 19 accounts banned; of these:
  * 7 of these were found to be spreading RU misinformation. Of these, 3 had additional focus, one was also found to be spreading hate against LGBTQ, 1 was found to also spread general hate and 1 was found to spread hate against immigrants and UA refugees); the other 4 were general Russian misinformation
    * 1 more specific ban focuses on videos concerning the Great replacement theory
    * The rest were banned due to various array of violation
  * 16 accounts were found not violating and therefore were not banned. The reason for this ranged from none or very low violations, through satirical accounts focused on politics, to openly pro-Russian accounts, which, however, have no violations at all.
  * The last 12 accounts were not banned but remain on our watchlist. This can be due to having a few violations. A couple of these are specifically Russian misinformation leaning.

Escalation 2

Deepfake of the leader of Progressive Slovakia - Michal Šimečka.

* Reported by CMS and immediate action from our side. We prevented reuploads and performed platform sweep to remove already published duplicates.

Escalation 3

* CMS have reported a number of videos still live on the platform, of which most were in breach of our CGs. Unfortunately, most were approved by the machine because they didn t reach a moderation threshold (not enough vvs, weren t reported, there are no visual violations, it is unlikely that any SK audio violation would have been caught by the machine, etc.), only one was mismoderated - we are following up on this now. These videos also include 3 already user deleted ones and 3 who are currently appealing for their bans.

Escalation 4

* Deepfake of Šimečka and Tódová. Tódová is a journalist and one of the subjects of LGBTQ hatred. Escalated by CMS among their video reports, once we saw what it was, the video was subject to the HRL library – prevents reuploads, platform sweep and content actioning.

Elections Content Trends

* Hate speech toward migrants

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_019056

- Degrading speech concerning how illegal migrants are dangerous to people and Slovak society as a whole; Documenting the arrival of illegal migrants oftentimes accompanied by derogatory remarks.
- *Trigger*: An increasing flow of migrants coming across the border from Hungary, mostly from Syria, to Slovakia. Small Slovak towns on the southern border were seeing some refugees rest or sleep over in their parks and sports grounds, which started discussions around citizens and border safety.

- Protecting "Traditional values" against LGBTQ+ people and gender ideology
  - Degrading speech concerning 1) the absurdity of the existence of more than two genders and the forced inclusive language (example: Progressive Slovakia party and their followers were attacked by using the term pregnant people instead of pregnant women); 2) the need to stop gender ideology from schools to protect minors; 3) LGBT described as plague (or other diseases) by politicians as well as users. Moreover, many videos showed clips/compilations from different Pride parades or other LGBTQ+-related events taken out of context, which position LGBTQ+ people as abnormal or even dangerous.
  - *Triggers*: 1) The terrorist attack in Bratislava on Tepláreň LGBT bar in October 2022 - two men died after shooting, which started a public conversation on anti-LGBT hate and LGBTQ+ rights. 2) Hungary passed a law (2021) banning LGBT content in schools or kids TV while outlawing sharing information deemed as promoting homosexuality with minors, which started a conversation about whether a similar law should be replicated in Slovakia. 3) The Progressive Slovakia party (*Progresívne Slovensko*) created a liberal political program which argues for giving LGBTQ+ people the same rights regarding marriage and adoption of children, which received backlash from many conservative parties and users.

- Election misinformation

  Misinformation concerning that:
  - Elections are about to / will be stolen (for example, Progressive Slovakia party or Eset, Slovak IT company, will manipulate election results);
  - Voters need to bring their pens because when they use pens from voting stations, their votes will be erased, and other parties/politicians will be selected;
  - The votes by mail (from Slovaks living abroad) will be manipulated or miscalculated.

- Deepfakes uploaded onto the platform close to election day
  - Čaputová, the Slovak president, stated that she is going to vote for the Republika party [redacted]

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

- Šimečka, the leader of the Progressive Slovakia party, is planning a drastic increase in the price of beer [redacted]
- Šimečka and Monika Tódová, the journalist, are planning election manipulation (ex. [redacted]

* Conspiracy theories related to George Soros
  - Zuzana Caputova [or another high-profile Slovak politician in these elections] is regularly paid by Soros, works directly for Soros or is Soros agent;
  - Claims mentioning George Soros + nefarious agendas; e.g. Soros wants to bring immigrants or LGBTQ+ agenda to destroy Europe/Slovakia, Christian culture/values, destabilise countries...

* Physical altercation depicting high-profile politicians
  - A physical fight between Matovič (former PM) and Kaliňák (ex-interior minister);
  - Tomáš Holkovič (OĽANO politician) punching a man who spat onto his face.
  - The footage of the fights remains on the platform as it was mainly posted to document/criticise the low standards of current Slovak politics. In relation to that, another trend emerged: Mortal Kombat editions (example here), which represented political satire (displaying different politicians as fighters in a PC game).

Monitization Integrity

* TikTok took a number of actions to mitigate the risk of political advertising in Slovakia for the election period, including building a war room, and implementing keywords related to political figures and misinformation trends in local languages as well as Russian and Ukrainian. TikTok saw no political ads on the market for September and the days following the election into October and the Monetisation Integrity team received no escalations for the Slovakian elections

Influence Operations

* Please reach out to [redacted] for any CIO disruption information during this period.

Public Policy [redacted]

Cooperation with Council for Media Services

TikTok started cooperation with Council for Media Services in February 2023. CMS has competencies in monitoring illegal content online and cooperating with online platforms according to Slovak Media Law no. 264/2022. The CMS was onboarded to TikTok CPC program in

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

March and subsequently sent various requests for removals as well as for additional information regarding our polices. InJune, GR facilitated the meeting between CMS and members of TikTok's Slovak Election Working Group. As a CPC program member, CMS obtained access to the TSET system in October. During the election period, GR maintained ongoing contact with CMS and handed over CMS requests and reports.

GPPPA accounts identification and verification

GR identified political accounts on the platform and conducted labeling process. All major political parties in Slovakia were contacted and informed about account verification procedure as well as on no-political adds TikTok policy.

Contacts with Slovak Interior Ministry

TikTok contacted Slovak Interior Ministry to agree on the best possible sources of authoritative electoral information for our users. We also extended an offer to join the educational and awareness raising campaign on election disinformation run by the Ministry.

Appendix

Election Hub

Redacted

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 212

Appointment

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]
Sent: 4/27/2023 5:59:34 AM
To: CNECT-COP-TASK-FORCE@ec.europa.eu; STUMP Krisztina [redacted]@ec.europa.eu]; RABBACHIN Alberto (CNECT) [redacted]@ec.europa.eu]; BRAIDA Mattia [redacted]@ec.europa.eu]; GILLET Jeanne [redacted]@ec.europa.eu]; BURIAN Peter [redacted]@ec.europa.eu]; BIRGER Albin [redacted]@ec.europa.eu]; BLASCHKE Yannic [redacted]@ec.europa.eu]; GILAR Simon [redacted]@ec.europa.eu]; [redacted] [redacted]@meta.com]; [redacted] [redacted]@google.com]; [redacted] [redacted]@google.com]; [redacted] [redacted]@google.com]; [redacted] [redacted]@twitter.com]; [redacted] [redacted]@microsoft.com]; [redacted] (Microsoft)' [redacted]@microsoft.com]; [redacted] [redacted]@microsoft.com]; [redacted] [redacted]@linkedin.com]; [redacted] [redacted]@linkedin.com]; [redacted] [redacted]@tiktok.com]; 'Katya Viadziorchyk (AI Forensics)' [redacted]@aiforensics.org]; 'Marc Faddoul' [redacted]@aiforensics.org]; Jelena Berkovic [redacted]@faktograf.hr]; Carlos Hernandez [redacted]@maldita.es]; Katarina Klingova [redacted]@globsec.org]; Jorge Gomes [redacted]@vosteurope.org]; Pawel Terpilowski [redacted]@demagog.org.pl]; Ines Narciso [redacted]@vosteurope.org]; Joao Ribeiro [redacted]@vosteurope.org]; Salvatore Romano [redacted]@aiforensics.org]; Heather Dannyelle Thompson [redacted]@democracy-reporting.org]; Michael Meyer [redacted]@democracy-reporting.org]; Dominika Hadju [redacted]@globsec.org]; Jana Kazaz [redacted]@globsec.org]; [redacted]@eeas.europa.eu; Stanislav Matejka [redacted]@rpms.sk]; Timea Cervenova [redacted]@rpms.sk]; Omri Preiss [redacted]@alliance4europe.eu]
CC: KOLODZIEJSKA Anna [redacted]@ec.europa.eu]; FILIPEK Jon [redacted]@ec.europa.eu]

Subject: Working Group on Elections
Location: Microsoft Teams Meeting

Start: 5/11/2023 5:30:00 AM
End: 5/11/2023 6:30:00 AM
Show Time As: Tentative

Recurrence: (none)

Produced to HJC

This Message Is From an External Sender

Dear all,

Thank you for expressing interest in joining the Working Group on Elections.

Please find below the connection details for the first meeting. As mentioned in our previous email, the aim of the meeting is to:

- officially launch the working group,
- designate a coordinator,
- define the working group's objectives and working strands

We are looking forward to getting started on this important work,

Kind regards,
The CoP disinfo team

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Microsoft Teams meeting

Join on your computer, mobile app or room device
Click here to join the meeting

Meeting ID: [redacted]
Passcode: [redacted]
Download Teams | Join on the web

Join with a video conferencing device
[redacted]@vteams.tech.ec.europa.eu
Video Conference ID: [redacted]
Alternate VTC instructions

Joining is available only from a Commission videoconferencing system.

Learn More | Help | Meeting options

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 213

Message

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]
Sent: 6/22/2023 3:29:37 AM
To: BIRGER Albin [redacted]@ec.europa.eu]; BLASCHKE Yannic [redacted]@ec.europa.eu]; BRAIDA Mattia [redacted]@ec.europa.eu]; BURIAN Peter [redacted]@ec.europa.eu]; Carlos Hernandez [redacted]@maldita.es]; [redacted] [redacted]@tiktok.com]; [redacted] [redacted]@microsoft.com]; [redacted] [redacted]@bytedance.com]; [redacted] [redacted]@google.com]; Dominika Hadju [redacted]@globsec.org]; FILIPEK Jon [redacted]@ec.europa.eu]; [redacted]@bytedance.com]; GILAR Simon [redacted]@ec.europa.eu]; GILLET Jeanne [redacted]@ec.europa.eu]; [redacted] [redacted]@linkedin.com]; Heather Dannyelle Thompson [redacted]@democracy-reporting.org]; Ines Narciso [redacted]@vosteurope.org]; Irene Larraz [redacted]@newtral.es]; Jana Kazaz [redacted]@globsec.org]; Jelena Berkovic [redacted]@faktograf.hr]; Joao Ribeiro [redacted]@vosteurope.org]; Jorge Gomes [redacted]@vosteurope.org]; Katarina Klingova [redacted]@globsec.org]; Katya Viadziorchyk [redacted]@aiforensics.org]; KOLODZIEJSKA Anna [Anna.KOLODZIEJSKA@ec.europa.eu]; [redacted] [redacted]@meta.com]; [redacted]@bytedance.com]; [redacted] [redacted]@google.com]; Marc Faddoul [redacted]@aiforensics.org]; [redacted] [redacted]@twitch.tv]; [redacted] [redacted]@microsoft.com]; [redacted] [redacted]@linkedin.com]; Michael Meyer [redacted]@democracy-reporting.org]; [redacted] [redacted]@google.com]; Omri Preiss [redacted]@disarm.foundation]; Pawel Terpilowski [redacted]@demagog.org.pl]; RABBACHIN Alberto (CNECT) [redacted]@ec.europa.eu]; Salvatore Romano [redacted]@aiforensics.org]; Sam Jeffers [redacted]@whotargets.me]; [redacted] [redacted]@bytedance.com]; Stanislav Matejka [redacted]@rpms.sk]; STUMP Krisztina [redacted]@ec.europa.eu]; [redacted]@eeas.europa.eu; Tímea Červeňová [redacted]@rpms.sk]; [redacted]@microsoft.com]; Tommaso Canetta [redacted]@pagellapolitica.it]; [redacted]@bytedance.com]
Subject: WG on Elections - Agenda and ToR

Produced to HJC

This Message Is From an External Sender

Dear Working Group's Members,

Please find below the agenda points of today's meeting (3:30-4:30 CET):

- Working Group administration (co-chairs appointment, frequency of meetings,...)
- Terms of Reference: start of the discussions
- AOB

We have uploaded a draft Terms of Reference on the Teams space (here), which will serve as a basis to start our discussion today.

Kind regards,

The CoP Disinfo team

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 214

CMS

Snap Elections in Slovakia

Situation and activities presented by the Slovak Council for Media Services

Redacted

CoP WG on Elections, July 2023

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Snap Elections in Slovakia

- Date: September 30, 2023
- Situation:
  - Increase in the occurrence of mis- and disinformation, as well as hate speech and content that may undermine the integrity of the elections.
  - Spread of false narratives in connection to domestic and foreign policy (US supremacy, Slovak president presented as US agent, pro-Russian narratives, the war in Ukraine).
  - The narrative on stolen elections has been presented since May 2023 by various parties with pro-Russian and authoritative tendencies.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Disinformation in Slovakia

- Longterm problem on major online platforms (the most used platform is Facebook - also widely used by politicians for political messaging)
- Fact-checking capacities are low.
  - No home-grown fact-checking organisation
  - Now - up to 2 Slovak fact-checkers in AFP and up to 4 in Demagog.cz (only for Meta platforms).
- In general, distrust of the population towards traditional media and the government. Population tends to incline to conspiracy theories and false narratives (rise during Covid-19 pandemic and now with the war in Ukraine).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Activities of the CMS

CMS

- National Regulatory Authority responsible for media oversight and enforcement of regulatory frameworks related to AVMS.
- New competencies related to the AVMSD regime and illegal content online.
- Competencies associated with online platforms:
  - research and analytical activities, particularly regarding the dissemination of hate speech, disinformation, content that may seriously impair the development of minors, cyberbullying, and media literacy,...
  - cooperating with online platforms in efforts to enforce their respective community standards or ToS,
  - specific illegal content online (definition stemming from the SK penal code).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

CMS's cooperation with online platforms

CMS

- Cooperation with the platforms in the effective, proportionate, and non-discriminatory application of their community rules, norms, and standards.
- Activities:
  - Monitoring online content (cooperation with Police Force, government bodies, and NGOs),
  - Analysis and application of the platforms' ToS,
  - Escalation of content violating the platforms' ToS (application of platforms' policies).
- Active cooperation with Meta, Google and TikTok.
- In June 2023, CMS held bilateral meetings with representatives of Meta, TikTok and Google/Youtube and presented a set of requests connected to upcoming elections.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Requests connected to upcoming elections in Slovakia

- Placement of **information panels with links to reliable and official sources of information** about the electoral process.
- Emphasis on **adhering to the blackout period** when publishing online political advertisements and ensuring transparency.
- Emphasis on fulfilling the **commitments arising from the Code of Practice on Disinformation**, with a focus on the transparency of political advertising.
- Collaboration between platforms and non-profit organizations on **media literacy campaigns** during the elections.
- The Council for Media Services asks platforms to cooperate and maintain intensive communication when evaluating reports. This includes prioritizing and effectively evaluating reports to platforms regarding the spread of disinformation narratives on election manipulation and cooperation with State Electoral Commission.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Outputs from meetings

Blackout period

- Meta and Google – the responsibility is on advertisers.
- TikTok – does not allow paid political advertising in Slovakia.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Next steps for CMS

CMS

- Before elections:
  - Active cooperation with platforms - collecting information on fulfilment of requests (e.g. launch of media literacy campaigns and information panels).
  - Monitoring of the transparency of political advertising and tools in ads libraries.
  - Monitoring of mis- and disinformation campaigns and narratives in cooperation with the Police Force, government bodies, and NGOs and flagging the content to online platforms.
- After the elections, CMS plans to publish a comprehensive report mapping activities of online platforms related to Slovak snap elections in 2023 (to be published in November 2023).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 215

Appointment

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]
Sent: 7/18/2023 3:32:54 AM
To: BIRGER Albin [redacted]@ec.europa.eu]; BLASCHKE Yannic [redacted]@ec.europa.eu]; BRAIDA Mattia [redacted]@ec.europa.eu]; BURIAN Peter [redacted]@ec.europa.eu]; Carlos Hernandez [redacted]@maldita.es]; [redacted] [redacted]@tiktok.com]; [redacted] [redacted]@microsoft.com]; [redacted] [redacted]@bytedance.com]; [redacted] [redacted]@google.com]; Dominika Hadju [redacted]@globsec.org]; FILIPEK Jon [redacted]@ec.europa.eu]; [redacted] [redacted]@bytedance.com]; GILAR Simon [redacted]@ec.europa.eu]; GILLET Jeanne [redacted]@ec.europa.eu]; [redacted] [redacted]@linkedin.com]; Heather Dannyelle Thompson [redacted]@democracy-reporting.org]; Ines Narciso [redacted]@vosteurope.org]; Irene Larraz [redacted]@newtral.es]; Jana Kazaz [redacted]@globsec.org]; Jelena Berkovic [redacted]@faktograf.hr]; Joao Ribeiro [redacted]@vosteurope.org]; Jorge Gomes [redacted]@vosteurope.org]; Katarina Klingova [redacted]@globsec.org]; Katya Viadziorchyk [redacted]@aiforensics.org]; KOLODZIEJSKA Anna [redacted]@ec.europa.eu]; [redacted] [redacted]@meta.com]; [redacted] [redacted]@bytedance.com]; [redacted] [redacted]@google.com]; Marc Faddoul [redacted]@aiforensics.org]; [redacted] [redacted]@twitch.tv]; [redacted] [redacted]@microsoft.com]; [redacted] [redacted]@linkedin.com]; Michael Meyer [redacted]@democracy-reporting.org]; [redacted] [redacted]@google.com]; Omri Preiss [redacted]@alliance4europe.eu]; Pawel Terpilowski [redacted]@demagog.org.pl]; RABBACHIN Alberto (CNECT) [redacted]@ec.europa.eu]; Salvatore Romano [redacted]@aiforensics.org]; Sam Jeffers [redacted]@whotargets.me]; [redacted] (google) [redacted]@google.com]; [redacted] [redacted]@bytedance.com]; Stanislav Matejka [redacted]@rpms.sk]; STUMP Krisztina [redacted]@ec.europa.eu]; [redacted]@eeas.europa.eu; Timea Cervenova [redacted]@rpms.sk]; [redacted] [redacted]@microsoft.com]; Tommaso Canetta [redacted]@pagellapolitica.it]; [redacted] [redacted]@bytedance.com]

Subject: WG elections - last meeting before summer
Attachments: image001.png
Location: Microsoft Teams Meeting

Start: 7/19/2023 6:00:00 AM
End: 7/19/2023 7:00:00 AM
Show Time As: Tentative

Recurrence: (none)

Produced to HJC

This Message Is From an External Sender

Agenda for Wednesday meeting:

- Update on coordinators
- Terms of reference for the working group (latest version discussed with co-coordinators available in the WG teams space WG Elections_ToR_2023_v4.docx)
- Slovakian elections preparedness
- After summer planning

Microsoft Teams meeting

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Join on your computer, mobile app or room device
Click here to join the meeting

Meeting ID: [redacted]
Passcode: [redacted]
Download Teams | Join on the web

Join with a video conferencing device
[redacted]@vteams.tech.ec.europa.eu
Video Conference ID: [redacted]
Alternate VTC instructions

Joining is available only from a Commission videoconferencing system.

Learn More | Help | Meeting options

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 216

Appointment

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]
Sent: 7/20/2023 7:10:18 AM
To: BIRGER Albin [redacted]@ec.europa.eu]; BLASCHKE Yannic [redacted]@ec.europa.eu]; BRAIDA Mattia [redacted]@ec.europa.eu]; BURIAN Peter [redacted]@ec.europa.eu]; Carlos Hernandez [redacted]@maldita.es]; [redacted] [redacted]@tiktok.com]; [redacted] [redacted]@microsoft.com]; [redacted] [redacted]@bytedance.com]; [redacted] [redacted]@google.com]; Dominika Hadju [redacted]@globsec.org]; FILIPEK Jon [redacted]@ec.europa.eu]; [redacted]@bytedance.com]; GILAR Simon [redacted]@ec.europa.eu]; GILLET Jeanne [redacted]@ec.europa.eu]; [redacted] [redacted]@linkedin.com]; Heather Dannyelle Thompson [redacted]@democracy-reporting.org]; Ines Narciso [redacted]@vosteurope.org]; Irene Larraz [redacted]@newtral.es]; Jana Kazaz [redacted]@globsec.org]; Jelena Berkovic [redacted]@faktograf.hr]; Joao Ribeiro [redacted]@vosteurope.org]; Jorge Gomes [redacted]@vosteurope.org]; Julian Merer [redacted]@arcom.fr]; Katarina Klingova [redacted]@globsec.org]; Katya Viadziorchyk [redacted]@aiforensics.org]; KOLODZIEJSKA Anna [redacted]@ec.europa.eu]; [redacted]@meta.com]; [redacted] [redacted]@bytedance.com]; [redacted]@google.com]; Marc Faddoul [redacted]@aiforensics.org]; [redacted] [redacted]@twitch.tv]; [redacted] [redacted]@microsoft.com]; [redacted] [redacted]@linkedin.com]; Michael Meyer [redacted]@democracy-reporting.org]; [redacted] [redacted]@google.com]; Nicolas Faucouit [redacted]@arcom.fr]; Omri Preiss [redacted]@alliance4europe.eu]; Pawel Terpilowski [redacted]@demagog.org.pl]; RABBACHIN Alberto (CNECT) [redacted]@ec.europa.eu]; Salvatore Romano [redacted]@aiforensics.org]; Sam Jeffers [redacted]@whotargets.me]; [redacted] (google) [redacted]@google.com]; [redacted] [redacted]@bytedance.com]; STUMP Krisztina [redacted]@ec.europa.eu]; [redacted]@eeas.europa.eu; [redacted] [redacted]@microsoft.com]; Tommaso Canetta [redacted]@pagellapolitica.it]; [redacted] [redacted]@bytedance.com]

Subject: WG elections - meeting on elections in Slovakia
Location: Microsoft Teams Meeting

Start: 9/5/2023 6:00:00 AM
End: 9/5/2023 7:00:00 AM
Show Time As: Tentative

Recurrence: (none)

Produced to HJC

This Message Is From an External Sender

Microsoft Teams meeting

Join on your computer, mobile app or room device
Click here to join the meeting

Meeting ID: [redacted]
Passcode: [redacted]
Download Teams | Join on the web

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Join with a video conferencing device

[redacted]@vteams.tech.ec.europa.eu
Video Conference ID: [redacted]
Alternate VTC instructions

Joining is available only from a Commission videoconferencing system.

Learn More | Help | Meeting options

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 217

Appointment

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]
Sent: 7/20/2023 7:12:17 AM
To: CNECT-COP-TASK-FORCE@ec.europa.eu; BIRGER Albin [redacted]@ec.europa.eu]; BLASCHKE Yannic [redacted]@ec.europa.eu]; BRAIDA Mattia [redacted]@ec.europa.eu]; BURIAN Peter [redacted]@ec.europa.eu]; Carlos Hernandez [redacted]@maldita.es]; [redacted]@tiktok.com]; [redacted]@microsoft.com]; [redacted]@bytedance.com]; [redacted]@google.com]; Dominika Hadju [redacted]@globsec.org]; FILIPEK Jon [redacted]@ec.europa.eu]; [redacted]@bytedance.com]; GILLET Jeanne [redacted]@ec.europa.eu]; [redacted]@linkedin.com]; Heather Dannyelle Thompson [redacted]@democracy-reporting.org]; Ines Narciso [redacted]@vosteurope.org]; Irene Larraz [redacted]@newtral.es]; Jana Kazaz [redacted]@globsec.org]; Jelena Berkovic [redacted]@faktograf.hr]; Joao Ribeiro [redacted]@vosteurope.org]; Jorge Gomes [redacted]@vosteurope.org]; Julian Merer [redacted]@arcom.fr]; Katarina Klingova [redacted]@globsec.org]; Katya Viadziorchyk [redacted]@aiforensics.org]; KOLODZIEJSKA Anna [redacted]@ec.europa.eu]; [redacted]@meta.com]; [redacted]@bytedance.com]; [redacted]@google.com]; Marc Faddoul [redacted]@aiforensics.org]; [redacted]@twitch.tv]; [redacted]@microsoft.com]; [redacted]@linkedin.com]; Michael Meyer [redacted]@democracy-reporting.org]; [redacted]@google.com]; Nicolas Faucouit [redacted]@arcom.fr]; Omri Preiss [redacted]@alliance4europe.eu]; Pawel Terpilowski [redacted]@demagog.org.pl]; RABBACHIN Alberto (CNECT) [redacted]@ec.europa.eu]; Salvatore Romano [redacted]@aiforensics.org]; Sam Jeffers [redacted]@whotargets.me]; [redacted] (google) [redacted]@google.com]; [redacted]@bytedance.com]; STUMP Krisztina [redacted]@ec.europa.eu]; [redacted]@eeas.europa.eu; [redacted]@microsoft.com]; Tommaso Canetta [redacted]@pagellapolitica.it]; [redacted]@bytedance.com]
CC: [redacted]@crispthinking.com]; [redacted]@feedback.org

Subject: WG elections - standing meeting
Attachments: image001.png
Location: Microsoft Teams Meeting

Start: 9/15/2023 6:00:00 AM
End: 9/15/2023 7:00:00 AM
Show Time As: Busy

Recurrence: (none)

Produced to HJC

This Message Is From an External Sender

Dear all, please see today's agenda

- finalise and adopt Terms of Reference (WG Elections_ToR_2023_v4.doc)
- discuss Risk assessment and rapid response system
- debrief of members on elections
- setup meeting focused on Slovakia

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Microsoft Teams meeting

Join on your computer, mobile app or room device
Click here to join the meeting

Meeting ID: [redacted]
Passcode: [redacted]
Download Teams | Join on the web

Join with a video conferencing device
[redacted]@vteams.tech.ec.europa.eu
Video Conference ID: [redacted]
Alternate VTC instructions

Joining is available only from a Commission videoconferencing system.

Learn More | Help | Meeting options

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 218

TikTok Technology Limited
Company Nr: 635755
Registered office: 10 Earlsfort Terrace,
Dublin, Ireland
D02 T380

European Commission,
1049 Brussels, Belgium
Redacted

Redacted

Dear Redacted and team,

Letter concerning the SK election

Thank you for your letter dated 5 September 2023, we have now had a chance to fully consider our response. We agree with the Commission that the upcoming Slovakian election requires a careful and considered risk management response and we look forward to providing further information about TikTok's measures.

TikTok's approach to election integrity in Slovakia

We are keen to continue the existing dialogue between the Commission and TikTok in relation to election integrity and the upcoming Slovakian election in particular. You will recall that we presented to you about our general measures in a meeting on 17 July 2023.

We also provided a non-paper detailing our specific measures for Slovakia on 10 August 2023. On 4 September 2023, we presented on our measures for Slovakia to the Election Group working group under the Code of Practice on Disinformation. We are pleased to submit an updated non-paper (see **enclosed**) with additional details of measures that we are now in a position to speak publicly about, consisting of:

- **Digital literacy partner.** We are engaging the well-respected digital literacy partner, DigiQ, to develop our media literacy campaign for Slovakian users, aimed at educating them on how to detect misinformation and verify information they are seeing online. DigiQ's educational videos will be available from our in-app Election hub, and we will redirect users to authoritative information on their website;

- **Trusted flaggers:** DigiQ and Slovakian Council for Media Services can escalate any potential violative content directly to us. We are also working hard to on-board more members of civil society to the programme ahead of the election; and

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

- **Speaker series.** We are engaging DigiQ to share their insights and market expertise with our internal teams to inform our approach to the upcoming election.

As well as the specific measures we are implementing for the Slovak election, our reports under the Code of Practice on Disinformation provide an overview of all of the measures that we implement for our EU community to combat disinformation on our platform, as well as the metrics underlying those measures. This information is provided at member state level, including for Slovakia. We reported at this level for our first COPD report in January and we will do so in our next report which will be published next week, on 15 September. You will recall that TikTok was the only platform to produce granular, country-level metrics in our first COPD report.

We note the Commission's request to provide further written explanations ahead of the meeting on Tuesday. Given the urgent timing of the meeting, we trust that the updated non-paper is sufficient to provide our initial written response to the questions numbered 1 to 6 as outlined on page 2 of your letter. We will be able to provide any further detailed required in the meeting or by way of a further written update, if needed.

Attendees from TikTok

I am pleased to set out below our attendees for the meeting on 12 September. These attendees represent the staff with the relevant technical and subject matter expertise, as they have an in depth understanding of our terms, polices and enforcement procedures. Our in person attendee, Redacted is already known to the Commissions DSA team and has a deep understanding of all T&S functions with over 20 years' experience in this field. He is also responsible for overseeing all elections programs within EMEA.

Name	Role	In person or virtual
Redacted	Redacted Middle East & Africa	In person
Redacted	Redacted	
Redacted	Redacted	Virtual
Redacted	Redacted	
Redacted	Redacted	
Redacted	Redacted Threat Disruption	
Redacted	Redacted Threat Disruption	
Redacted	Redacted Political and Regulatory Risk	
Redacted	Redacted Content Regulation	

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Our suggested agenda

Thank you for bringing together a diverse group of attendees for the session on Tuesday. In order to make the best use of the valuable time together, we suggest the following agenda and speakers:

- Introductions - all
- Overview of the TikTok service in Slovakia - Redacted
- Overview of the authorities' expectations for the meeting and sharing insights - Commission and Authorities
- Our specific election integrity measures for Slovakia- Redacted and others as needed
- Overview of our approach to political advertising on TikTok- Redacted
- Our content moderation approach to election misinformation (with reference to our non-paper)- Redacted
- Our approach to detection of covert influencing Redacted Redacted
- Q&A

Yours sincerely

Redacted

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 219

Message

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]
Sent: 9/1/2023 7:31:13 AM
To: [redacted]@tiktok.com]; [redacted]@microsoft.com]; [redacted]@microsoft.com]; [redacted]@google.com]; [redacted]@google.com]; [redacted]@linkedin.com]; [redacted]@meta.com]; [redacted]@bytedance.com]; [redacted]@google.com]; [redacted] (Microsoft)' [redacted]@microsoft.com]; [redacted]@linkedin.com]
CC: 'Jana Kazaz' [redacted]@globsec.org]; Tommaso Canetta - PagellaPolitica [redacted]@pagellapolitica.it]; Carnapete, Louise [redacted]@eui.eu]; Dominika Hadju [redacted]@globsec.org]; 'Michael Meyer' [redacted]@democracy-reporting.org]; 'Heather Dannyelle Thompson' [redacted]@democracy-reporting.org]; [redacted]@tiktok.com]; STUMP Krisztina [redacted]@ec.europa.eu]; RABBACHIN Alberto (CNECT) [redacted]@ec.europa.eu]; GILLET Jeanne [redacted]@ec.europa.eu]; BRAIDA Mattia [redacted]@ec.europa.eu]
Subject: FW: Follow up on WG elections last meeting

This Message Is From an External Sender

Dear all,

A gentle reminder of the below request to share your planned measures, actions and envisaged cooperations regarding Slovakian elections as today is the deadline the WG agreed on.

This will be then shared on the teams space of the WG ahead of the meeting on Tuesday, 5 September.

Regards,

The CoP team on behalf of the WG coordinators

Produced to HJC

From: CNECT COP TASK FORCE
Sent: Thursday, July 20, 2023 10:38 AM
To: BIRGER Albin (CNECT) <[redacted]@ec.europa.eu>; BLASCHKE Yannic (CNECT) <[redacted]@ec.europa.eu>; BRAIDA Mattia (CNECT) <[redacted]@ec.europa.eu>; BURIAN Peter (CNECT) <[redacted]@ec.europa.eu>; 'Carlos Hernandez' [redacted]@maldita.es>; [redacted]' <[redacted]@tiktok.com>; [redacted]' <[redacted]@microsoft.com>; [redacted]' <[redacted]@bytedance.com>; [redacted]' <[redacted]@google.com>; 'Dominika Hadju' <[redacted]@globsec.org>; FILIPEK Jon (CNECT) <J[redacted]@ec.europa.eu>; [redacted]' <[redacted]@bytedance.com>; GILAR Simon (CNECT) <[redacted]@ec.europa.eu>; GILLET Jeanne (CNECT) [redacted]@ec.europa.eu>; [redacted]' <[redacted]@linkedin.com>; 'Heather Dannyelle Thompson' <[redacted]@democracy-reporting.org>; 'Ines Narciso' <[redacted]@vosteurope.org>; 'Irene Larraz' <[redacted]@newtral.es>; 'Jana Kazaz' [redacted]@globsec.org>; 'Jelena Berkovic' <[redacted]@faktograf.hr>; 'Joao Ribeiro' [redacted]@vosteurope.org>; 'Jorge Gomes' <[redacted]@vosteurope.org>; 'Julian Merer' <J[redacted]@arcom.fr>; 'Katarina Klingova' <[redacted]@globsec.org>; 'Katya Viadziorchyk'

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

<[redacted]@aiforensics.org>; KOLODZIEJSKA Anna (CNECT) <[redacted]@ec.europa.eu>; [redacted]
<[redacted]@meta.com>; [redacted]' [redacted]@bytedance.com>; [redacted]'
[redacted]@google.com>; 'Marc Faddoul' <[redacted]@aiforensics.org>; [redacted]' [redacted]@twitch.tv>; '[redacted]
[redacted] <[redacted]@microsoft.com>; [redacted]' <[redacted]@linkedin.com>; 'Michael Meyer'
<[redacted]@democracy-reporting.org>; [redacted]' <[redacted]@google.com>; 'Nicolas Faucouit'
<[redacted]@arcom.fr>; 'Omri Preiss' <[redacted]@alliance4europe.eu>; 'Pawel Terpilowski'
<[redacted]@demagog.org.pl>; RABBACHIN Alberto (CNECT) <[redacted]@ec.europa.eu>; 'Salvatore
Romano' <[redacted]@aiforensics.org>; 'Sam Jeffers' <[redacted]@whotargets.me>; [redacted] (google)'
<[redacted]@google.com>; [redacted]' <[redacted]@bytedance.com>; STUMP Krisztina (CNECT)
<[redacted]@ec.europa.eu>; SYKORA Ondrej (EEAS) <[redacted]@eeas.europa.eu>; [redacted]'
<[redacted]@microsoft.com>; 'Tommaso Canetta' <[redacted]@pagellapolitica.it>; [redacted]'
[redacted]@bytedance.com>

Subject: Follow up on WG elections last meeting

Dear Members of the WG on elections,

As discussed during yesterday's meeting, please see below the relevant points to follow up:

1. Terms of Reference

Please find below the link to the Terms of Reference documents.

WG Elections_ToR_2023_v4.docx (/Crisis subgroup/WG on elections)

The group addressed the comments during the meeting and identified solutions to address them. Those who could not attend the end of the meeting, **please indicate if you have any major concerns regarding the draft as it stands - by 26 July COB at the latest** - and we will get in touch together with the co-coordinators to discuss. The aim is to close the document before the summer break with only formal confirmation needed in September.

2. Preparedness for Slovakia

We encourage Members to share documents and actions they are taking with regards to the upcoming elections on Slovakia by email to the WG members during the summer period.

Fact-checkers and Civil Society Organisation Members of the WG have shared the importance to exchange proactively ahead of the elections in Slovakia. They indicated that they would like to understand from platforms **what actions they are planning and what measures they are putting in place.** Importantly, **they would like to know in particular if the platforms are interested in planning ad hoc initiatives ahead of this election** (eg. showing pre-bunking content to their users that are looking at related content, specifically tailored on the disinformation that was viral in Slovakia in the past months, for example about Ukrainian refugees?), and **what these initiatives could eventually be so that relevant cooperations can be established with fact-checkers and relevant NGOs**.

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

The Commission supports this initiative and **invites platform signatories to share their planned measures, actions and envisaged cooperations as soon as possible, but in any event before the 1 September.**

A **dedicated meeting on election preparations for the Slovakian elections will take place at the beginning of September, indicatively 5 September, 15:00**. Please indicate if you are interested in joining this meeting.

3. Next meeting

The **next meeting of the full Working group is planned for 14 September**, we will follow up with an invite.

Happy summer to everyone!

The CoP team

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 220

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]
Sent: 9/1/2023 2:31:13 PM
To: Redacted@tiktok.com]; Redacted microsoft.com]; Redacted Redacted microsoft.com]; Redacted google.com]; Redacted google.com]; Redacted linkedin.com]; Redacted meta.com]; Redacted [Redacted bytedance.com]; Redacted google.com]; Redacted [Redacted microsoft.com]; Redacted linkedin.com]
CC: Redacted globsec.org]; Redacted pagellapolitica.it]; Redacted Redacted eui.eu]; Redacted globsec.org]; Redacted [Redacted democracy-reporting.org]; Redacted democracy-reporting.org]; Redacted tiktok.com]; Redacted @ec.europa.eu]; Redacted Redacted @ec.europa.eu]; Redacted @ec.europa.eu]; Redacted Redacted ec.europa.eu]
Subject: [External] FW: Follow up on WG elections last meeting

Dear all,

A gentle reminder of the below request to share your planned measures, actions and envisaged cooperations regarding Slovakian elections as today is the deadline the WG agreed on.

This will be then shared on the teams space of the WG ahead of the meeting on Tuesday, 5 September.

Regards,

The CoP team on behalf of the WG coordinators

From: CNECT COP TASK FORCE
Sent: Thursday, July 20, 2023 10:38 AM
To: Redacted ec.europa.eu>; Redacted <Redacted ec.europa.eu>; Redacted ec.europa.eu>; Redacted Redacted @ec.europa.eu>; Redacted maldita.es>; Redacted <Redacted tiktok.com>; Redacted microsoft.com>; Redacted <Redacted bytedance.com>; Redacted google.com>; Redacted <Redacted globsec.org>; Redacted ec.europa.eu>; Redacted <Redacted bytedance.com>; Redacted ec.europa.eu>; Redacted Redacted @ec.europa.eu>; Redacted linkedin.com>; Redacted Redacted democracy-reporting.org>; Redacted vosteurope.org>; Redacted Redacted newtral.es>; Redacted globsec.org>; Redacted <Redacted faktograf.hr>; Redacted vosteurope.org>; Redacted <Redacted vosteurope.org>; Redacted arcom.fr>; Redacted <Redacted globsec.org>; Redacted aiforensics.org>; Redacted <Redacted ec.europa.eu>; Redacted meta.com>; Redacted <Redacted bytedance.com>; Redacted google.com>; Redacted <Redacted aiforensics.org>; Redacted twitch.tv>; Redacted <Redacted microsoft.com>; Redacted linkedin.com>; Redacted <Redacted democracy-reporting.org>; Redacted google.com>; Redacted

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

<[Redacted]arcom.fr>; [Redacted]alliance4europe.eu>; [Redacted]
<[Redacted]demagog.org.pl>; [Redacted]@ec.europa.eu>;
[Redacted]aiforensics.org>; [Redacted]whotargets.me>; [Redacted]
<[Redacted]google.com>; [Redacted]bytedance.com>; [Redacted]
[Redacted]@ec.europa.eu>; [Redacted]eeas.europa.eu>; [Redacted]
<[Redacted]microsoft.com>; [Redacted]pagellapolitica.it>; [Redacted]
<[Redacted]bytedance.com>

Subject: Follow up on WG elections last meeting

Dear Members of the WG on elections,

As discussed during yesterday's meeting, please see below the relevant points to follow up:

1. Terms of Reference

Please find below the link to the Terms of Reference documents.

WG Elections_ToR_2023_v4.docx (/Crisis subgroup/WG on elections)

The group addressed the comments during the meeting and identified solutions to address them. Those who could not attend the end of the meeting, **please indicate if you have any major concerns regarding the draft as it stands - by 26 July COB at the latest** - and we will get in touch together with the co-coordinators to discuss. The aim is to close the document before the summer break with only formal confirmation needed in September.

2. Preparedness for Slovakia

We encourage Members to share documents and actions they are taking with regards to the upcoming elections on Slovakia by email to the WG members during the summer period.

Fact-checkers and Civil Society Organisation Members of the WG have shared the importance to exchange proactively ahead of the elections in Slovakia. They indicated that they would like to understand from platforms **what actions they are planning and what measures they are putting in place.** Importantly, **they would like to know in particular if the platforms are interested in planning ad hoc initiatives ahead of this election** (eg. showing pre-bunking content to their users that are looking at related content, specifically tailored on the disinformation that was viral in Slovakia in the past months, for example about Ukrainian refugees?), and **what these initiatives could eventually be so that relevant cooperations can be established with fact-checkers and relevant NGOs**.

The Commission supports this initiative and **invites platform signatories to share their planned measures, actions and envisaged cooperations as soon as possible, but in any event before the 1 September.**

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

A **dedicated meeting on election preparations for the Slovakian elections will take place at the beginning of September, indicatively 5 September, 15:00**. Please indicate if you are interested in joining this meeting.

3. Next meeting

The **next meeting of the full Working group is planned for 14 September**, we will follow up with an invite.

Happy summer to everyone!

The CoP team

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 221

Ref. Ares(2023)6026012 - 05/09/2023

EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR COMMUNICATIONS NETWORKS, CONTENT AND TECHNOLOGY

Platforms Policy and Enforcement
Digital Services

Brussels
5 September 2023

Redacted
Via email Redacted@bytedance.com

Dear Madam,

TikTok was designated as a very large online platform through Commission Decision C(2023) 2720 final of 25 April 2023. Pursuant to Article 34(1) of Regulation (EU) 2022/20651 (hereafter "DSA"), providers of very large online platforms and of very large online search engines shall carry out risk assessments by the date of application referred to in Article 33(6) DSA, second subparagraph 2. The Commission services also drew your attention to this obligation in the letters of 11 August and 1 September 2023.

Article 34 and Article 35 DSA oblige very large online platforms (VLOPs) and very large online search engines (VLOSEs) to identify, analyse and assess any actual or foreseeable negative effects on democratic processes, civic discourse and electoral processes, as well as public security stemming from their services, and to put in place reasonable, proportionate, and effective mitigation measures. Insofar as relevant mitigation measures take the form of restrictions in your terms and conditions, Article 14 DSA obliges VLOPs and VLOSEs to act in a diligent, objective and proportionate manner in applying and enforcing such restrictions. By the time this letter was sent, the Commission services had not received reports setting out the results of the risk assessment pursuant to Article 34 DSA from you.

As stated in the European Democracy Action Plan, the rapid growth of online campaigning and online platforms has also opened new vulnerabilities for democracies and made it more difficult to maintain the integrity of elections, ensure a free and plural media, and protect the democratic process from disinformation and other manipulation. Given the importance of protecting our democracies, the Commission services therefore put strong emphasis on the category of risks in the DSA that concerns the actual or foreseeable negative effects on democratic processes, civic discourse and electoral processes, as well as public security.

The 2023 Slovak parliamentary election, which are to be held on 30 September 2023, are the first national election in the EU, that take place after the DSA has become legally enforceable for VLOPs and VLOSEs. Your service is considered to be an important platform for public discourse during these upcoming elections. Therefore, pursuant to the

Commission européenne/Europese Commissie, 1049 Bruxelles/Brussel, BELGIQUE/BELGIË – Tel. Redacted
Office: L-51 04/DCS – Tel. direct line Redacted

CNECT-DSA-Registry@ec.europa.eu

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

powers under Article 72 DSA, the Commission services request technically detailed explanations on your risk assessment, mitigation measures, and enforcement capacities for your terms and conditions regarding this election. These explanations should be given orally in the context of a meeting on 12 September 2023 with relevant Slovak authorities in Bratislava. In addition, the Commission services strongly encourage you to submit contextualising written information in advance of the meeting. To facilitate a constructive meeting, we request the presence of staff with relevant technical expertise and in-depth understanding of your enforcement processes and capacities for your terms and conditions in the relevant areas.

These explanations should cover at least the following items:

1. The risk assessment process for prioritising and allocating resources to different elections to ensure the effectiveness of your election integrity efforts.
2. The outcome of this process for the Slovak election, including the reasoning and evidence substantiating your decision-making process. If applicable, the insights and outcome of any red teaming exercise you might have conducted in preparation for the election.
3. The measures put in place to mitigate the identified risks in the context of the Slovak election, including performance metrics you use to assess the effectiveness of these mitigation measures.
4. In case these mitigation measures take the form of restrictions on content or user behaviour on your service via your terms and conditions:
    a. Performance metrics you have set internally for your efforts to detect infringements of the specific restrictions in Slovakia.
    b. Your detection and enforcement capacities for the specific restriction in Slovakia, including team structures, their respective human resources, and linguistic capacities.
    c. Temporary changes to these enforcement capacities resulting from your risk assessment for the Slovak election.
5. The structure of your cooperation with trusted partners from civil society and research that might have relevant real-time information for your efforts to mitigate risks regarding the Slovak election.
6. The structure of your cooperation with industry partners to share threat intelligence on influence operations and coordinated attempts to manipulate your service.

In case you decide to send written information in advance of or as a follow-up to this meeting, please use the following e-mail address: CNECT-DSA-Registry@ec.europa.eu. If specified, the information received shall be treated as DSA confidential information and thus will not be shared with anyone outside the Commission services.

Please also send a list of participants, their functions with the company and whether you would like to attend the meeting in person or in a hybrid manner to this email address. The Commission services further ask you to send an acknowledgement of receipt within 24 hours from the receipt of this letter to the same email address.

The Commission services are available to answer any questions you may have related to this topic.

Yours faithfully,

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted

c.c.: Redacted @bytedance.com; Redacted @tiktok.com; Redacted @tiktok.com

Electronically signed on 05/09/2023 14:27 (UTC+02) in accordance with Article 11 of Commission Decision (EU) 2021/2121

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 222

[TnS - PL] Parliamentary Elections - 2023 - Election Moderation Guidelines

Drafted by	[redacted]
Sign Off	☑ ~~Election integrity~~ [redacted] ☑ ~~Harmful Misinformation~~ [redacted] ☑ ~~Hateful Behaviour/Harassment~~ [redacted] ☑ ~~VSDA~~ [redacted] ☑ ~~Shocking and Graphic Content~~ [redacted] ☑ ~~EMEA Regional Policy~~ [redacted]
Status	Sign off
Valid from	09/10/2023
Scope	PL market

Background

Parliamentary elections will be held in Poland on **15 October 2023** to elect members of the Sejm and Senate, which will consist of 460 members of parliament and 100 senators for a four-year term. The previous election, in 2019, resulted in the re-election of a majority Law and Justice government, albeit without a majority in the Senate. The Law and Justice government will try to win a record 3rd term

Additional info: A referendum is scheduled to be held concurrently, touching on the following topics:

- approval of privatisation of state-owned enterprises;
- increase in the retirement age;
- admission of immigrants under the EU relocation mechanism;
- removal of the barrier on Poland's border with Belarus

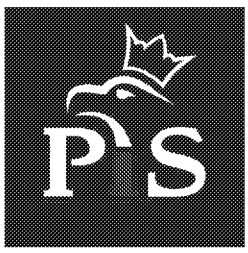
Voting process takes place only on the 15th October 2023 7 a.m. - 9 p.m. local time

Polish nationals of at least 18 years old are eligible to vote.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

The election is organised by the National Electoral Commission (Państwowa Komisja Wyborcza - PKW)

Parties and Candidates

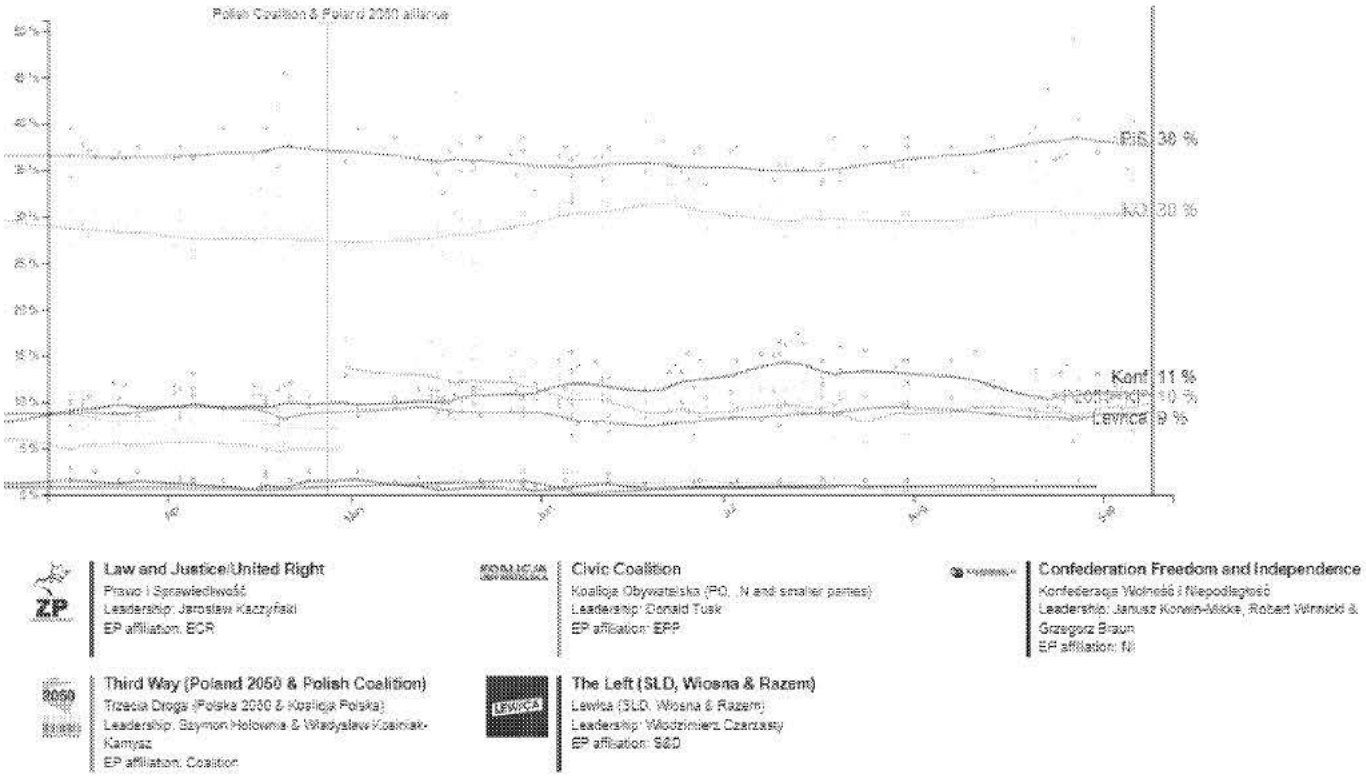
Party / Candidates	Related visuals	Notes
Prawo i Sprawiedliwość (Law and Justice/United Right) Leader: Jarosław Kaczyński		The **nationalist Law and Justice Party**, which took power in 2015 and was victorious in the 2019 election, is seeking a third term. The government's tenure has been marred with clashes with the European Union over the government's rule of law record and democratic backsliding. It is known for a harder-line stance regarding immigration, LGBT rights, women's rights, and the role of religion in the country. The party is campaigning on promises of new spending on social and military causes.
Koalicja Obywatelska (Civic Coalition) Leader: Donald Tusk		The **main opposition competitor Civic Coalition** is campaigning on vows to reverse the negative trends in foreign and domestic policy, as well as mending fences with Brussels to secure funds frozen by the EU amid the rule of law dispute. Party leader Donald Tusk, who is a former prime minister and former top EU figure, also promised to free state media and cultural activities from what the coalition calls their current restrictions and "censorship."

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Konfederacja Wolność i Niepodległość (Confederation Freedom and Independence) Leaders: Sławomir Mentzen and Krzysztof Bosak	
Trzecia Droga (Third Way) Leaders: Szymon Hołownia & Władysław Kosiniak-Kamysz	
Lewica (The Left) Leaders: Włodzimierz Czarzasty and Robert Biedroń	

- Other major parties running include **Confederation Freedom and Independence** (whose increase in popularity is attributed to social media, including TiKTok)**, Third Way, and The Left**.
- **Opinion polls** suggest that the right-wing Law and Justice grouping, Poland's governing party since autumn 2015, will emerge as the largest party but fall short of an overall majority. It is currently averaging 36-37% support but needs around 40% to have a chance of securing an unprecedented third term in office.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED



Graphic source: politico.eu

- **Campaigning On Hate Speech:** This is the third election season where the ruling party has made use of hateful messaging as part of its campaigning tools. For example, in 2019, LGBT+ rights were described by the party's chairman, Jarosław Kaczyński, as a "threat to Polish identity," as it sought to maximise the conservative Catholic vote. Migration is another topic used by the party. Recently, Prime Minister Mateusz Morawiecki stated that the rape of a young Polish man in Munich, allegedly by an Afghan refugee, was a consequence of Europe's "open borders."
  - Other major topics that come up during the pre-election campaigning include women's rights (abortion), the country's relations with Russia, considerations around Russian war in Ukraine and refugees.

Policy Instructions and Case Examples

📌
- **Please pay attention to any content mentioning elections in captions, hashtags, audio and video (text overlay included).**
- **If in doubt whether a policy would apply, please err on the side of caution and tag the policy.**
- **Multi-tagging rules apply unless stated otherwise.**

1. Integrity and Authenticity

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_019082

1.1 Election Misinformation

We remove misleading content specifically related to participation in an election or other civic events. Some election misinformation examples may include:

Scope [Tag Policy]	Examples	Screenshot/Video
False or misleading information about how to vote	• "You can vote from your phone" • "You can only vote if you are 21 years old" • "The day/time for voting has changed" • False information about voting hours • "Make sure to mark all the parties in the ballot whom you do not support" • "You need to provide vaccination proof to vote" • "You can't vote if you don't speak the language" • "You can vote with your driver's license" • "You need to take a photo of your vote to prove you voted" • "You can vote through this app" • "You need to bring your own pen to vote" • A candidate claims early victory AND encourages people to stop participating in the election	
Misinformation about laws, processes, and procedures that govern the organization and implementation of elections and other civic processes	• "If you vote blank, your vote will be added to the majority vote" • "If you need assistance to cast your vote, only a family member can assist you"	

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	• "The pen used in poll stations can be easily erased" • Claims that the polls have closed early, when in reality the voting process continues. • "The Ministry of Affairs is the institution that organizes the elections"	**Redacted**
		• "Change in elections: You can vote for 2 PiS party members on a single ballot"
False or misleading information about how to register to vote	• "You can only register to vote if you are +25" • "The deadline for registering has passed" • "You can't register if you don't have your passport"	
False or misleading information about qualifications to run in an election	• "X candidate can't run because there is a corruption investigation against her"	
False claims about technical eligibility requirements for candidates and officials to run for office	• "You need to be 35 years old to run for Senate" • "You need to have a high school diploma to be a candidate"	
False claims about the outcome of an election, once the results are final.	• "Duda is not the President of Poland" • "The final results are not official yet" • "We can change the final results if we sign this petition"	

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Not in scope [Do Not Tag]
Public interest exceptions (documentary, counter speech, satire and humorous content)
Calls for vote "Null" or "Blank" (form of political speech)
Calls for abstaining from voting (form of political speech)
Video inside polling stations (even if local regulations restrict or disallow video/photography in election booths)
Content depicting filled or unfilled ballot cards
Content depicting a party's election programme in a misleading way
Suggesting a candidate might win, before official results are published
Content related to vote profiteering or vote buying
Content related to public opinion polling

1.2 Dangerous Misinformation

We remove false or misleading content that is likely to cause panic and could result in real world harm to a person, group, organization or institution.

Scope [Tag Policy]	Examples
Content which makes misleading claims and encourages people to destroy public property, before, during and after Election Day	• "Go and destroy those ballot boxes because PiS is manipulating elections" • "Steal those electronic voting machines before Konfederacja can hack them"
Content re-purposing old videos, falsely claiming that the event or video is new in order to trigger societal panic and fear during the election	• Footage of past terrorist or violent attacks in polling stations, repurposed as current events • Footage of past criminal gang's activities in polling stations, repurposed as current events • Footage of past attacks against candidates, politicians, voters or poll workers, repurposed as current events • Footage of past military activity, repurposed as current events
False information warning of an alleged catastrophe, epidemic, or other emergency which encourages violence against individuals or groups (e.g. Political opponents) during the election	• False claims of a terrorist attack at a polling station

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_019085

	• "Elections were postponed because of the earthquake. They are preparing themselves to steal the election, we need to stop this" • False claims that the military is preparing to attack the population during the election • False claims regarding obstruction to electoral registration that calls for retaliatory violence • False claims based on fake documents supposedly issued by government officials warning about an imminent event that may cause panic and could result in real world harm • Content falsely claiming that a politician or candidate has died
Content falsely claiming basic necessities (food, water, etc) or services (banks, cash machines, etc.) are no longer available in a particular location, causing hoarding or over purchasing during the elections	• "ATMs will not be available on Election Day" • "Public transportation will be suspended/unavailable/interrupted during Election Day"
Not in scope [Do Not Tag]	
• Public interest exceptions (documentary, counter speech) • Non-deceptive or satirical depiction of politicians' look-alikes	

1.3 Dangerous Conspiracy Theory

We remove highly specific conspiracy theories that directly target a living person(s), attack a specific protected group, include a violent call to action, or deny a violent or tragic event create [sic] a significant amount of harm.

Scope [Tag policy]	Examples	Screenshot/Video
Content that is (i) divisive and polarizing AND (ii) dispute historical events and/or generally held belief/fact AND (iii) assign blame to a living person(s) by name.	• False theories that link candidates or political campaigns to terrorist organizations	

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Conspiracies that completely deny that a violent event/tragedy happened	• Conspiracies that deny abuses under dictatorships, such as assassinations and enforced disappearance • Conspiracies denying the loss of lives in violent tragedies.	
Any negative conspiratorial content that targets any protected groups.	• "Ukraine was promised EU membership if it could overthrow the Polish government"	Redacted
Conspirational content that is accompanied by a violent call to action, and/or has previously been accompanied by a violent call to action	• "They stole the election, let's close Sejm"	

1.4 Fake News

We remove content that is packaged in a way where the user would believe the information is coming from a credible news source when it's not.

Scope [Tag policy]	Examples	Screenshot/Video
Misleading content that is presented as news, or coming from a credible news source, when it is not		

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	• This video contains false news and is packaged to look like it is from a legitimate and trustworthy news site, when it's not	Redacted
	• The paper is real, but taken out of context - blurred part on the right-hand side + the entire interview	
Misleading screenshots or images of a news article that have been modified	• Front cover of The Economist is doctored, the title of the cover has been modified.	
Misleading video content with a news source banner, or with someone who appears like a professional reporter, when they are not	• This video shows someone who appears to be a professional reporter, but is not. It uses a real news outlet banner.	

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Not in scope [Do Not Tag]		
• Credible content by a news source • Satire, humorous content (please check if it's a parody account before actioning on the content) • News content debunking fake news • Misleading content NOT presented as news or coming from a news source	• Criticising Donald Tusk in a non-violating way/Anti-Tusk propaganda which does not meet the threshold of a conspiracy theory or fake news	**Redacted**

1.5 Conspiracy Theories Related to Current, Unfolding Events

We control for theories that directly relate to a current and unfolding event where details and facts are still developing and can cause moderate to significant public harm by increasing fear, panic and confusion about what is true and what is false.

Note: Depending on the market, moderators can be instructed to tag Election Misinformation in order to move unverified content to the IO queue for further review.

Scope[Tag policy]	**Examples**	**Screenshot/Video**
Call into question the nature of present day events where details and facts are still developing and/or not confirmed.	This includes unverified electoral claims, such as: • Unverified claims that state a candidate has died suddenly or become ill • The electoral authority's website goes down on Election Day, without a verified explanation, and content provides conspirational explanations	

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	• The electoral authority takes more time than usual to provide preliminary electoral results, without a verified explanation, and content provides conspirational explanations • Unverified claims that a violent event is happening in an electoral precinct • Candidate/user/media outlet declares a candidate's early victory before an official announcement • Claims that a certain group or organization is responsible for a violent or tragic event, when the details have not been verified.	Redacted Claim that Ukrainian government is threatening Poland

1.6 General Conspiracy Theory

We control conspiracy theories that are not directed at a specific person(s), but target an organization or institution and create a moderate amount of harm, potentially leading to distrust in the targeted institution.

Scope[Tag policy]	Examples	Screenshot/Video
Content that calls into question a generally held belief or fact by questioning what, why or how something happened, AND suggests an institution could be responsible.	• "Candidates are financed/influenced by Russia" [When this is ascertained not to be the case] • "The government is trying to demobilize voters by using covid lockdowns" [When this is ascertained not to be the case]	

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

- A fire at the Election Commission is part of a plan to alter electronic voting machines [When this is ascertained not to be the case]

Redacted

- Claims that Tusk was ready to give up a big part of Poland to Russia in 2011

Not in scope [Do Not Tag]

- Content describing unexplained events/phenomena/myths

1.7 Impersonation (Escalate to PQA/CPM)

Please escalate accounts potentially impersonating politicians/candidates/political parties

Scope	Example
• Accounts impersonating politicians/political parties/candidates (impersonating info in handle/bio/avatar) **Please escalate to PQA/CPM**	Redacted

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted

Not in scope	Counterexample
• Partisan political accounts/fan accounts with misleading/official sounding bios • Satire/fan accounts clearly marked as such • Parody or commentary accounts clearly marked as such	Fan account Redacted

2. Hateful Behaviour

2.1 Hate Speech

Scope[Tag policy]	Examples	Screenshot/Video
Hate speech is defined as speech that does or intends to attack, threaten, incite violence against, or dehumanize an individual or a group of individuals on the basis of protected attributes.	"Scientists announced that after a man dolphins are the most intelligent species. That means women dropped to third place"	Redacted

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Not in scope [Do Not Tag]		
Content which does not contain direct attack against a Protected Group, but critising/discussing migration policies	'In Germany, the Afghans and the Pakistani commit to rape 16 times more often than the Germans. What do you think would happen if we allowed illegal immigrants to our country? (...) Polish women feel 48 times safer than those in GB. (...) Let's not waste what we have achieved. Let's think about our safety.' 'Most criminal activity is reported in countries that are widely open to immigrants. According to Eurostat data, the most dangerous countries in Europe are Spain and Belgium. 96% of Poles, consider their area as safe and 88% of people consider Poland a safe country. In all published statistics, Poland is mentioned as one of the safest places, especially for women.' Caption: "MEP Iratxe García Pérez respond to misogynistic comments made by Polish MEP Janusz Korwin-Mikke (2017) #womensliberation #misogyny" Video should not be tagged as hate speech since the user is focusing on the response to misogynic remarks.	**Redacted** **Redacted**

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

2.2 Use of a slur in a degrading or hateful context

Scope[Tag policy]	Examples	Screenshot/Video
Slur usage in political speech should still be tagged	Slur "cwel" at 00:20	Redacted
Not in scope [Do Not Tag]		

2.3 Promotion of Hateful Ideologies

Scope[Tag policy]	Examples	Screenshot/Video
Promotion of a hateful ideology includes: ・Speaking positively or glorying a hateful ideology or hate figure (white supremacy, male supremacy...) ・Songs or lyrics promoting a hateful ideology		

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

· Displaying symbols associated with a hateful ideology		Redacted
Not in scope [Do Not Tag]		

3. Bullying and Harassment

Policy Reminder

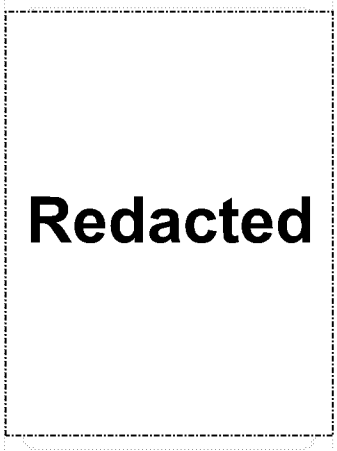
Please be reminded that political candidates in the context of campaigning are considered public figures. Therefore, please keep in mind that *Bullying Statements* and *Bullying Behaviors* policies have exceptions when the target is a public figure, unless the attack is a "severe attack".

Exception when the target is a public figure	No exception when the target is a public figure
• Bullying Statements • Bullying Behaviors	• Sexual Harassment • Revealing or Threatening to Reveal Others's Personal Data / Personally Identifiable Information • Praising or Belittling Violent Tragedies and their Victims

4. Shocking and Graphic Content

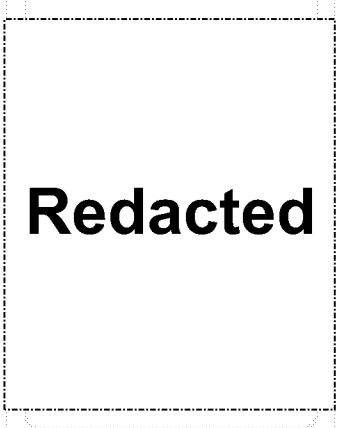
4.1 Violent Death or Accident in a News or Fictional Setting

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Scope[Tag policy]	Examples	Screenshot/Video
Content shared as a part of political discourse and depicting socially/politically significant events	0:05 - violent accident with a person	

5. Violent Extremism

5.1 Promotion of Terrorism (International)

Scope[Tag policy]	Examples	Screenshot/Video
Any non-state actors using premeditated violence or threats of violence to cause harm to non-combatant individuals, in order to intimidate or threaten a population, government, or international organization in the pursuit of political, religious, ethnic, or ideological objectives.	Depiction of designated VE individuals without clear EDSA context	 Theodore "Ted" Kaczynski

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 223

Appointment

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]
Sent: 9/20/2023 12:38:55 AM
To: CNECT-COP-TASK-FORCE@ec.europa.eu; [redacted] [redacted]@meta.com]

Subject: FW: WG on Elections - Slovak Elections Update
Location: Microsoft Teams Meeting

Start: 9/22/2023 5:30:00 AM
End: 9/22/2023 7:00:00 AM
Show Time As: Free

Recurrence: (none)

This Message Is From an External Sender

Produced to HJC

Dear WG Members,

As discussed, next meeting is dedicated to signatories' actions and updates on Slovak elections.

Best,
The CoP team

Microsoft Teams meeting

Join on your computer, mobile app or room device
Click here to join the meeting

Meeting ID: [redacted]
Passcode: [redacted]
Download Teams | Join on the web

Join with a video conferencing device
[redacted]@vteams.tech.ec.europa.eu
Video Conference ID: [redacted]
Alternate VTC instructions

Joining is available only from a Commission videoconferencing system.

Learn More | Help | Meeting options

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 224

[TnS - SK] Slovak snap election - 2023 - Election Moderation Guidelines

Status	Signed off
Notice Type	Trending Topic
Market(s)	CZ/SK
Effective Date	22 September 2023
End Date	N/A
Product Type(s)	Video
Round(s)	R1 / R2
Drafted by	[redacted]
Consulted	• HHB [redacted] • Harmful misinformation [redacted] • OPS - CZ/SK SME [redacted]
Sign-off by	☐ Harmful Misinformation [redacted] ☐ HHB [redacted] ☐ Civic Integrity [redacted] ☐ EMEA Regional Policy Lead [redacted]

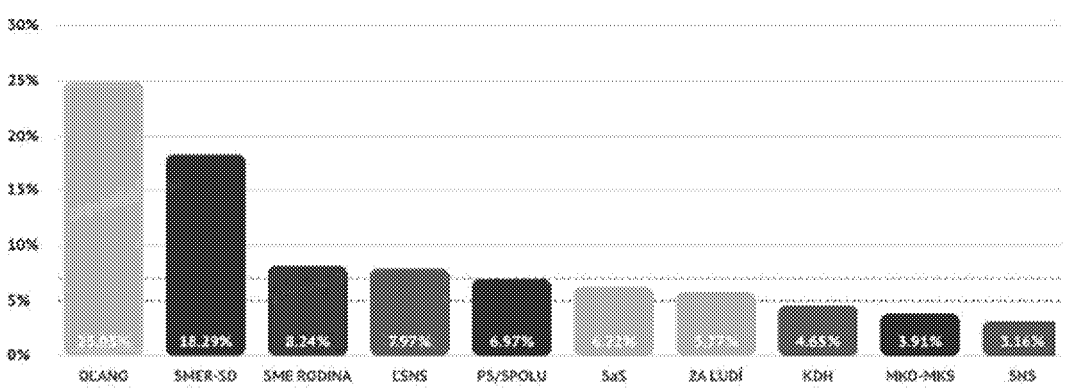
The snap election date: Saturday, September 30, 2023

This document provides the following:

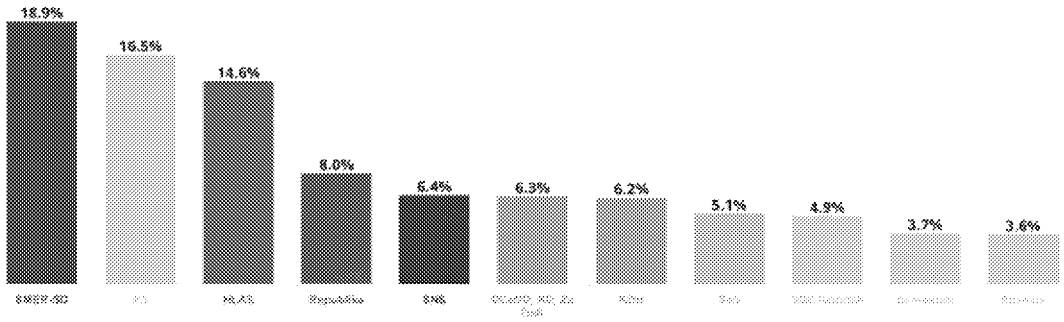
- Background info about the Slovak parliamentary elections and the main political parties/candidates.
- Policy instructions, reminders and case banks in relation to:
  - Hateful behaviour & harassment - focus: Hate towards LGBTQ+ community and migrants.
  - Integrity & Authenticity policies - focus: claims related to George Soros.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Background

Type of election	Parliamentary elections • Elections to the national parliament - National Council of the Slovak Republic (*Národná rada Slovenskej republiky*) • 150 members of the parliament are directly elected by universal adult suffrage for a 4-year term office • The winner of the election - the political party or coalition with the highest number of votes - is entrusted with the formation of the government. If the winner does not have the necessary support to gain the majority in the parliament (at least 76 deputies), another party/coalition can form the government.
Who can vote	A citizen of the Slovak Republic who turns 18 on the day of the election at the latest
When to vote	Saturday, September 30, 2023 between 7am - 10pm
How to vote	The voters need to: • Bring their identity card to the polling station. If people vote in a precinct in which they are not registered, they must show a voting card. • Choose one ballot with the name of one political party they want to vote for • Circle a maximum of 4 candidates (ordinal numbers are circled)
Voting from abroad	Voters from abroad must have applied for a vote via a mail by August 9, 2023. Those who registered received a special envelope and instructions on how to download ballots from the website of the Ministry of the Interior. The envelope with the vote must then be received by the authorities by September 29, 12pm (noon).
The last election	 Results of the parliamentary elections in 2020
Polls	

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED



Source: Fokus agency / SME (from 17.9.2023)

Main topics of elections

- Fighting poverty / Increasing the standards of living
- Enhancing health care system
- Support for Ukraine
- The change in Immigration policies / Protecting national borders
  - Trigger: An increasing flow of migrants coming across the border from Hungary, mostly from Syria. Small Slovak towns on the southern border are seeing more and more refugees rest or sleep over in their parks and sports grounds, which started a conversation around citizens and border safety.
- Protecting "Traditional values" vs. giving LGBTQ+ people equal rights & gender ideology - triggers:
  - The terrorist attack in Bratislava on Tepláreň LGBT bar in October 2022 - two men died after shooting, which started a public conversation on anti-LGBT hate and LGBTQ+ rights
  - Hungary passed a law (2021) banning LGBT content in schools or kids TV while outlawing sharing information deemed as promoting homosexuality with minors, which started a conversation about whether a similar law should be replicated in Slovakia
  - The Progressive Slovakia party (*Progresívne Slovensko*) created a liberal political program which argues for giving LGBTQ+ people the same rights regarding marriage and adoption of children, which received backlash from many conservative parties

Parties and Candidates

In total, there are 24 political parties in these elections.

The list below provides an overview of the 9 most popular political parties.

Political Party - Info	Brief Party Description	The most notable candidates

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

SMER
SLOVENSKÁ SOCIÁLNA DEMOKRACIA

- *SK:* Smer – sociálna demokracia
- *ENG:* Direction – Social Democracy

TikTok account:

Redacted

- Centre-left-wing populist
- Pro-Kremlin
- Associated with several corruption scandals

Core topics:

- Strict immigration policy
- Fighting against poverty
- Against multiculturalism – Advocating for traditional values

Redacted

Robert Fico (lead)

- Former Prime Minister (2006-2010 & 2012-2018)
- Resigned in 2018 in the wake of the political crisis following the murder of journalist Ján Kuciak

Redacted

Ľuboš Blaha

- Member of the SKparliament; political scientist
- Criticized for spreading hoaxes, conspiracy theories and disinformation; strongly pro-Kremlin
- Blocked by Meta since June 2022

Redacted

Róbert Kaliňák

- Former Minister of the Interior - the longest serving minister in the history of modern Slovakia

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Progresívne Slovensko

- *SK:* Progresívne Slovensko
- *ENG:* Progressive Slovakia

TikTok account:

Redacted

- Progressive & social-liberal
- Pro-EU and green

Core topics:

- Improving the Slovak economy and healthcare system
- Change of social atmosphere (so people stop leaving Slovakia)
- Advocating for the rights of LGBT+ people (e.g. same-sex unions); promoting gender equality and the religious neutrality of the state

Redacted

Michal Šimečka (lead)

- Vice-President of the European Parliament;
- The son of journalist Martin Milan Šimečka and a grandson of philosopher and dissident Milan Šimečka

Redacted

Michal Truban

- Businessman, founder of several startups;
  Former chairman of the party
- Criticized for talking to students about his experiences with marijuana *(note: Cannabis in SK is illegal and possession of even small amounts of the drug can lead to lengthy prison terms)*

SK: HLAS – sociálna demokracia

- Having a pro-European outlook
- Reflecting a more conservative stance on social issues
- Described as a "catch-all party" (reluctant to take positions opposed by the majority of voters)

Redacted

Peter Pellegrini (lead)

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

ENG: VOICE - Social democracy

TikTok account:

Redacted

- Former Prime Minister (2018-2020) and member of SMER
- Considered the most trustworthy Slovak politician (Jun, 2022; Feb 2023)

- SK: Republika
- ENG: Republic

Tiktok account:

Redacted

- Far-right; patriotic
- Often regarded as extremist and neo-fascist *(the party refuses such labels)*
- Eurosceptic, Russophilic, anti-American
- Strong anti-vaccination position during the COVID-19 pandemic
- In June 2023, YouTube terminated channels linked to the party
- Known for the spread of Russian wartime propaganda, anti-vaccination & anti-LGBTQ+ rhetoric

Core topics - Advocating for:

- Protection of traditional & Christian values
- Improvement the standard of living
- Against sending military aid to UA
- Complete suspension of government support to all NGOs, (calling them *"rainbow and Soros orgs)*

Redacted

Milan Uhrík (lead)

- Elected a member of the European Parliament in 2019 for the far-right neo-Nazi LSNS party. Later, he founded his own party, Republic.

Redacted

Milan Mazurek

- In 2019, he was convicted of defamation of race based on his anti-Romani statements, and became the first Slovak parliamentarian to lose his seat because of a crime

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

- SK: Obyčajní ľudia a nezávislé osobnosti
- ENG: Ordinary people and independent figures

- A center-right multi-party alliance
- Advocates strict immigration policies, traditional values and strong nationalism
- Pro-EU; Against multiculturism

Core topics:

- Anti-corruption & Anti-mafia

Redacted

Igor Matovič

- Former Prime Minister (2020-2021); businessman
- The most distrusted Slovak political figure in the Slovak public, with a 91% distrust score
- Many controversies - admitting plagiarizing his masters thesis; blamed for causing the current government crisis; recent physical fight with Smer s politicians...

Redacted

Veronika Remišová

- Also broadly distrusted Slovak politician (over 80% distrust score)
- Former Minister of Investments, Regional Development and Informatization

- SK: Sloboda a Solidarita
- ENG: Freedom and Solidarity

Centrist-liberal; Pro-EU

Core topics:

- Less government intervention in markets/economy

Redacted

Richard Sulík (Lead)

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

- SK: Slovenská národná strana
- ENG: Slovak National Party

Nationalist; Eurosceptic

- Former allegations of racism, discrimination and neo-fascism

Core topics:

- Protecting traditional values & Slovak culture
- Emphasing the Christian foundation of the state

Redacted

Andrej Danko (Lead)

- Controversies: Plagiarism allegation (doctoral thesis); Promotion to the rank of Captain - single promotion by 8 ranks (which was seen as a sign of corruption)

KĐH

- SK: Kresťanskodemokratické hnutie
- ENG: Christian Democratic Movement

Centre-right; pro-EU;

Christian-democratic

Expressing strong anti-LGBT sentiment

Redacted

Milan Majerský (Lead)

- Controversy: He recently labelled the LGBT community as a plague and compared it to corruption

SME RODINA

- SK: Sme Rodina
- ENG: We Are Family

National-conservative; right-wing populist

Redacted

Boris Kollár (Lead)

- Controversies: He has 13 children with 11 different women (often mocked on TikTok); Allegations that he forced other women who got pregnant from him to have abortions

Policy Instructions and Case Examples

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_019104

- Please pay attention to any content mentioning elections in captions, hashtags, audio and video (text overlay included).
- Multi-tagging rules apply unless stated otherwise in Opus.
- If in doubt whether a policy will apply for the content posted by high-profile politicians or parties, please send the case to your QAs/SMEs who will further discuss it with the policy team.

Please note the guidance below is not exhaustive and content must still be considered under all relevant titles and their respective PBRs.

1. Harassment & Hateful Behavior

Hate speech & Hateful behaviour

We do not offer protection for public figures to engage in hate speech under the guise of political expression . If political speech has the likelihood to cause harm or incite violence, it will be removed from the TikTok platform.

Within the Slovak parliamentary elections, hateful behaviour appears mainly in relation to the LGBTQ+ community and migrants. Therefore, case banks predominantly focus on hate in these areas.

1. Hate Speech

Hate speech fosters an atmosphere of exclusion on the platform and increases the risk of violence against protected groups by attacking or dehumanising them.

Any speech that violates our HHB guidelines must be accompanied by some framing that refutes the hateful statement or acknowledges that it is harmful.

- E.g. A news account uploads a snippet of a politician s speech that contains a hateful statement with the caption, '*President attacked trans youth.*' -- This would be considered a documentary/raising awareness exception [No tag]
- E.g. A user uploads a snippet from a TV political discussion where a politician says that LGBTQ people are the threat while denouncing such opinion in captions or within the video - "*I cannot believe what that politician said*!" [No tag]

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_019105

If the users post snippets from political discussions that contain multiple political views while some are hateful and others denounce such hateful statements, we would generally approve (e.g. neutrally posting a TV debate between multiple politicians on gay rights where one politician attacks gay people and another politician criticizes him for it).

Reminder - Protected groups

The exhaustive list of protected groups:

* Race (e.g. Asian), Ethnicity (e.g. Turk), National origin (e.g. French), Religion (incl. Atheism), Immigration status (e.g. migrants, asylum seekers) Sex (male, female; relating to biological differences), Sexual orientation (e.g. heterosexuality, bisexuality), Gender (male, female; relating to social constructs), Gender identity (e.g. transgender), Serious Illness or Disability (e.g. HIV, Diabetes, psychotic illnesses).

The following categories are not considered protected groups *(non-exhaustive list)*:	Not Hate speech	Hate speech We consider countries or religions (as an institution/ideology) to be a protected attribute when they are used as a proxy and attacked in the following ways: wishing death, murder, torture, kidnapping, stalking, rape, blackmail, suffocation and other violent attacks.
* States	Examples: * *Slovakia is the most disgusting country* * *I hate Russia*	* *Death to Russia* * *Let s set Ukraine on fire* 🔥
* Religions (as an institution/ideology) or religious texts	* *Islam is dangerous* * *Christianity is full of bullshit and hate* * *I agree that Swedish people should continue burning Koran*	* *Death to Islamists* [Proxy for Muslims, which are a protected group]. * *Join us to destroy synagogues* [Proxy for Jews]
* Political parties or political groups	* *Let s destroy right wing voters* * *Progressives are the biggest threat to Slovakia*	*N/A*

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	• *Lefties should be deported from this country* • *Liberalism is a hateful ideology*	

General instructions

Scope	Examples
• Explicit/implicit statement made by a political representative/candidate expressing hate against a protected group - including violent, dehumanising or intolerant speech • Explicit/implicit spreading of hate against individual politicians on the basis of the politician s membership of a protected group	• *Female politicians are really all just whores - that s how they got their positions.* • *Since Zuzana Čaputová is a woman, she does not belong to Bratislava castle but to the kitchen!* • *We should ban stupid Hungarians in our politics.*
Not in scope	**Counterexamples**
• Burning a national flag • Burning a political party s flag • Expressing disgust towards a political party/political ideology • Expressing criticism of governmental policies • Counterspeech to hate speech • Criticism of immigration policy based on economic/security reasons rather than based on hostility against a protected group • Criticism and arguments on policies related to gender or protecting traditional values without hostility against individuals with protected attributes	

Instructions regarding gender identity

Scope	Examples

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

• Dehumanizing content that claims transgender people are morally or mentally inferior.	Trans people are • Perverts, sexual predators, pedophiles, or groomers • Mistakes, moral degenerates, evil, or satanic • Mentally ill, or that their gender identity is a mental illness • Molest children at higher rates than cis people • Trying to groom children to "make" them transgender or promoting sex to children as part of a trans-rights campaign
• Denying the existence of transgender people	• Children cannot be trans • Trans people are lying • Trans women aren t real women
• Expression of aggressive insults, disgust or abnormality	• Look at the trans in this video: 😕😕😕😕" • Trans people are disgusting/not normal
• LGBTQIA+ people are trying to corrupt or sexualise children through education on gender identity or expression or sexual orientation	Includes referring to LGBTQIA+ inclusive school materials as sex abuse or attempts to sexualise children In the context of LGBTQ/gender ideology • We need to stop sexualisation of young people/children • We don t allow sexualisation of young people/children
• Transgender (or gender ideology) was invented by pedophiles	
Another policy • Targeted misgendering/deadnaming Intentionally misgendering others in a degrading manner or referring to a transgender individual by their former name (deadnaming) – tagged under Bullying statements instead	Example • Pointing to a photo of a transgender woman: HE is not a woman. = targeted misgendering (against a particular individual)

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Not in scope	Counterexamples
• Respectful conversations about transgender athlete participation ◦ Inc. arguments on the basis of biological or physiological differences	• Trans women should not compete alongside cis women ◦ Men are on average faster and stronger than women ◦ If they transitioned after puberty, trans women s ability is closer to men than cis women

Common political claims

! All the claims below are common in the Slovak political discussions. They were categorised as No tag. When we see those claims standalone or without additional hateful behaviour signals, we should not tag Hate speech. Still, they represent soft signals that we should watch out for.

Claim	Categorisation	Additional instructions/notes/context
• We don t want mass migration • We want to stop mass migration • We don t want any illegal migration	No tag	
• More illegal migrants may pose a security risk • We should reduce the number of illegal immigrants arriving from Hungary • To counter migrants crossing illegally, we need to strengthen border controls However, please note (!) • We hate illegal migrants = *Hate Speech*	No tag *[Hate Speech]*	• Criticism/arguments on immigration policy/process without hostility against migrants
• We refuse / want to stop gender ideology However, please note (!)	No tag *[Hate Speech]*	Gender ideology (or gender theory) consists of a variety of issues whilst not having a coherent definition. Thus, it is considered only as a soft signal.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

• We refuse / want to stop LGBTQ+ ideology or transgender ideology = *Hate Speech*		
• We refuse / want to stop globalism	No tag	
• There are only two genders SK: Sú len dve pohlavia However, please note (!) • There are only two genders SK: Sú len dva rody / dve rodové identity = *Hate speech*	No tag *[Hate Speech]*	In the Slovak language, the word 'pohlavie' is used for both sex as well as gender. Therefore, without other hard signals (e.g. denying/degrading other gender identities - SK: *rodové identity*), we would not tag any HHB policy for the '*Sú len dve pohlavia*' statement, since claiming there are only two sexes is not a violation.
• I don t agree that there are 17 (or 70) genders SK: Nesúhlasím, že je 17 (or 70) pohlaví	No tag	• Questioning how many genders there are without denying non-binary gender identities is not considered Hate speech • Slovak politicians usually use the number 17 or 70 (but it is applicable to other higher numbers too)
• We don t want a rainbow ideology • We want to stop a rainbow ideology SK: *dúhová ideológia*	No tag	• Some politicians use the word rainbow ideology to refer to liberal political parties or causes.
• We need to protect Christianity/Christian values from gender ideology	No tag	
• Marriage is between a man and a woman • Marriage is a unique bond between one man and one woman • Marriage should only be allowed between a man and a woman SK: *Manželstvo je zväzok muža a ženy / Manželstvo tvorí muž a žena / Manželstvo by malo byť dovolené len pre muža a ženu* (...)	No tag	Since this is the legal reality in the market (it is written in the Slovak constitution), we will allow factual statements which assert that marriage is between a man and a woman. However, if there are any other signs of hateful or degrading speech against a protected group, we should tag the policy.
• We can not allow same-sex marriage	No tag	
• We can not allow same-sex couples to adopt children	No tag	However, if such statements would be accompanied by stating that gay parents

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

		are not fit/suitable to have children, then we would tag Hate speech.
• We don t want to build special toilets for trans people	No tag	
• We need to protect traditional values / Christian values / traditional family model	No tag	

Case bank

LGBTQ+

Case	Hateful claims + description	Tag
Redacted	Hateful claims (SK) • Dúhový pride nie je nič iné iba obyčajný pochod a stretnutie úchylákov a deviantov. • LGBTI je nebezpečná ideológia • Liberáli chodia našim deťom vymývať mozgy. Hovoria im, že pohlavie sa dá zmeniť, že muž sa môže stať ženou, a žena sa môže stať mužom. • Ochránime deti pred LGBTI (...) (ENG) • Rainbow Pride is nothing but an ordinary march and meeting of deviants • LGBTI is a dangerous ideology • Liberals brainwash our children. They Redacted Redacted changed, that a man can become a woman, and a woman can become a man. • We will protect children from LGBTI (...) Aggressive insults (e.g. deviants) and positioning the LGBTQ community as a threat to children/society	Hate Speech
	Hateful claims (SK)	Hate Speech

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted

* (Maďarsko) schválilo zákon, ktorý chráni deti pred vplyvom pedofilov a násilnej sexuálnej agendy
* (Zákon) chce zamedziť styku pedofilov s mladou generáciou

(ENG)

* (Hungary) approved a law that protects children from the influence of pedophiles and a violent sexual agenda
* (The law) wants to prevent contact between pedophiles and the younger generation

Positioning the LGBTQ community as paedophiles; Calling the LGBTQ agenda a perversion; Mentioning sexualisation and corruption of children in the context of LGBTQ.

Hateful claims
(SK)

Hate Speech

* Pomýlencov, ktorí sa prezliekajú do ženských šiat, vojna by ich naučila

(ENG)

* The war would teach delusional (men), who dress up women s clothes

Expressing strong hostility toward the LGBTQ community (trans people) by implying that they need to be
re-educated by the war

Hateful claims
(SK)

Hate Speech

* [Overlaying text] LGBTI, transgender a gender ideológia a jej hrozba pre Slovensko
* Myslím si, že LGBTI ideológia, gender ideológia , transgender ideológia sú veľkou hrozbou pre Slovensko, rovnako ako korupcia. (...) Je to pliaga.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted	(ENG) • [Overlaying text] "LGBTI, transgender and gender ideology and its threat to Slovakia • [In the video] I think that LGBTI ideology, gender ideology, transgender ideology are a big threat to Slovakia, just like corruption. (...) It s a plague." Calling LGBTI & transgender ideology a disease (plague) and a threat	
Redacted ENG: No helicopters, we only accept male and female!	Hateful claims (SK) • Kým sa progresívny Mišo rozhodne či je dnes chlapec, dievča alebo vrtuľník, pre nás je rodová ideológia na školách neprijateľná. (ENG) • Until the progressive Mišo decides whether he is a boy, a girl or a helicopter today, gender ideology in schools is unacceptable for us. Dehumanising speech - comparing diverse gender identities to inanimate objects & trivialising the importance of gender expression	Hate Speech
	Hateful claims (SK) • [Overlaying text] Rodová ideológia je pliaga.	Hate Speech

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted	• Nepoznám LGBT a tie ďalšie písmenká ideológiu. Ja poznám rodovú ideológiu, tu pokladám aj ja za pliagu. Chcem proti nej bojovať. Vymyslenú pedofilmi (...)." (ENG) • [Overlaying text] Gender ideology is a plague. • I don t know LGBT and those other letters of ideology. I know gender ideology, I also consider it a plague. I want to fight against it. Invented by pedophiles (...)." The politician connects gender ideology to pedophilia	
	Showing a person with a non-binary gender expression/identity together with the underlying text Sick times while the video is edited in a way to belittle the video subject	Hate Speech
	Purposely choosing/editing footage from pro-LGBTQ protests that position LGBTQ people as dangerous/abnormal	Hate Speech

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted		
Counterexamples		
Redacted	Even though the beginning of the video contains hatespeech (footage of politician Majersky saying that LGBTI is a plague that can destroy Slovakia), it is clearly condemned later by Šimečka. In the video entirety, it is clear counterspeech.	No tag
	The first part of the video shows a female politician from Progresívne Slovensko who argues for same-sex marriages and thus, the change of the Slovak constitution. Later, Boris Kollár disagrees with her and refuses such a change, but he does it without attacking protected groups.	No tag

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted	The politician reads the Slovak constitution, including that marriage is a unique bond between a man and a woman. She further states that she hopes the constitution will stay the same and how many MP votes would be needed for constitutional change.	No tag
	The video shows the lead of the conservative OLANO party, Matovič, who criticizes Progresívne Slovensko political programme. As this video appears in the context of discussing party programmes/policies and does not directly attack people from protected groups, it should be allowed. (SK)	No tag

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted

* *(from 1:34s)* Nedovolíme šírenie LGBT propagandy na školách a manželstvá homosexuálov a transsexuálov, adopcií detí homosexuálmi a transsexuálmi. Nepodporíme zmenu pohlavia mladistvých, nedovolíme zmenu pohlavia na počkanie bez akýchkoľvek psychologického či psychiatrického vyšetrenia. Toto navrhuje Progresívne Slovensko v svojom programe. Kedykoľek, ktokoľvek príde na matriku že on sa cíti alebo ona už niečím zo 72 rôznych pohlaví, tak musí to akceptovať žena na matrike, lebo zmena pohlavia bude už len na základe sebaurčenia. Keď niekto príde že je jeden deň muž a druhý žena, tretí deň ono, štvrtý deň neviem čo, lebo má 72 možností, tak budú na matrike zapisovať ľudia pohlavie podľa toho ako sa kto cíti. (...) Hlúposť, zvrátenosť. (...) Toto nie je progres, toto je regres, toto je úpadok, niečo čo na normálne Slovensko nepatrí.

(ENG)

* *(from 1:34s)* We will not allow the spread of LGBT propaganda in schools and marriages of homosexuals and transsexuals, adoption of children by homosexuals and transsexuals. We will not support the change of gender of minors, we will not allow gender change without any psychological or psychiatric examination. This is what Progressive Slovakia proposes in its program. Whenever, whoever comes to the registry office that he or she already feels like one of the 72 different genders, the woman at the registry office must accept it, because the gender change will only be based on self-determination. When someone comes that one day is a man and the next a woman, the third day that, the fourth day I don t know what, because there are 72 options, then people will write down the gender on the register according to how they feel. (...) Stupidity, perversion. (...) This is not progress, this is regression, this is decline, something that does not belong to normal Slovakia.

Note: Matovič also calls other liberal policies proposed by Progresívne Slovensko to be stupid/perversions (concerning condoms, contraception, abortions, adoptions...).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted	The speaker on the grounds of the European parliament tries to argue against transwomen in sport by drawing an analogy with Mike Tyson joining the female boxing category. Respectful conversations about transgender athlete participation	No tag

Immigrants

Case	Description	Tag
Redacted	Hateful claims: (SK) * [Overlaying text] Budú behať vrahovia detí aj po našich detských ihriskách? * Koľko takýchto vrahov malých detí si budeme aj my musieť zobrať na Slovensko? Ak takýchto vrahov odmietneme prijať, budeme musieť za každého jedného platiť finančnú pokutu. * My si cez hranice týchto nelegálnych vrahov a násilníkov púšťať nebudeme. (...) Pretože Slovenská republika o takého formy kultúrneho obohatenia, o takéto formy terorizmu a kriminality rozhodne nestojí. * Zabránime tomu, aby slovenské deti museli zažívať to, čo zažívajú deti vo Francúzsku, vo Švédsku, v Nemecku a v iných štátoch, ktoré urobili túto katastrofálnu chybu a nechali sa zaplaviť takýmito sadistickými vrahmi a teroristami. (ENG)	Hate speech

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	* [Overlaying text] Will the killers of children also run around our playgrounds? * How many such murderers of small children will we have to take to Slovakia? If we refuse to accept such murderers, we will have to pay a financial fine for each one. * We will not let these illegal murderers and thugs cross our borders. (...) Because the Slovak Republic is definitely not interested in such forms of cultural enrichment, such forms of terrorism and crime. * We will prevent Slovak children from having to experience what children in France, Sweden, Germany and other countries are experiencing, which made this catastrophic mistake and allowed themselves to be overwhelmed by such sadistic murderers and terrorists. Attributing criminals behaviour to the group - migrants	
Redacted	Only sharing the news article about the rape in Plzeň would not hit our policies. However, the user further frames the article (see the text above the video), which makes it hate speech. Redacted ENG: How many Czechs need to die because of Ukrainians? How many raped women before you finally realise? Attributing criminals behaviour to the group - Ukrainian migrants	Hate speech
	(SK) Deti migrantov v Nemecku sa postupne začleňujú do výuky na školách. (ENG)	Hate speech

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted	Migrant children in Germany are gradually being integrated into school education. Showing a single incident from supposed-to-be class in Germany, while attributing such unwanted behaviour to all children of immigrants in Germany. = Attributing criminals behaviour to the group.	
	The video compiles different footage of immigrants on the streets combined with supposed-to-be empty houses/districts that local people do not want to occupy. The user purposely selected specific snippets to create an image that immigrants and local people cannot coexist. = Exclusion or intolerance toward immigrants	Hate speech
Counter-examples		
	Documenting the arrival of migrants in Velký Krtíš. There is no additional framing/speech by the user that would attack migrants.	No tag

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted		
Redacted	Non-violative claims: (SK) [Overlaying text] SLOVENSKO čelí masovému prílevu nelegálnych cudzincov! * Či sú naše hranice pred nelegálnymi a potenciálne nebezpečnými imigrantmi chránené o to sa skrátka nezaujímajú [referencia k terajšej vláde] * Zastavíme tento prílev nelegálnych imigrantov cez naše hranice * Nemajú tu čo robiť, porušujú naše zákony, sú potenciálnym ohrozením pre našich obyvateľov, a nám nesmieme záležať natom, aby počet nelegálnych imigrantov, ktorí na Slovensko budú v budúcnosti prichádzať, bola jedna veľká nula. (ENG) [Overlaying text] SLOVAKIA faces a mass influx of illegal foreigners! * Whether our borders are protected from illegal and potentially dangerous immigrants, they simply do not care [reference to the current government]	No tag

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	* We will stop this flow of illegal immigrants across our borders * They have no business here, they violate our laws, they are a potential threat to our residents, and we are extremely concerned that the number of illegal immigrants who will come to Slovakia in the future is one big zero.	
	Criticism and arguments on immigration policy based on security reasons rather than based on hostility against immigrants	
Redacted	Politician Kaliňák criticises the current government s approach to migration politics, while he mentions the case of Sweden, where the relaxed migration politics led to building ghettos where Swedish laws do not apply. [The debate happened in the context of the ongoing political discussion on how Slovakia should solve illegal immigration that comes from the Hungarian border.]	No tag
	Criticism and arguments on immigration policy based on security reasons rather than based on hostility against immigrants	

1.2 Promotion of Hateful Ideologies

General instructions

Scope	Examples
Representation, glorification, and promotion of Hateful Ideologies or figures	Ideologies likely to be found in content related to the elections: * Remigration * Great Replacement Theory/Grand Remplacement * Identitarianism * Muslims want to impose sharia law

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	• Referring to migrants/foreigners as invaders • Holocaust denial and other antisemitic conspiracies

Case	Description	Tag
Redacted	The video connects the contemporary footage of illegal immigrants coming to Slovakia with the Kalergi Plan & Great replacement theories, which suggest that there is a plot to mix/replace white Europeans with other races.	Promotion of hateful ideologies [Multi-tagged with Dangerous Conspiracy Theory]
Counterexamples		
Redacted	The viral video (created by using AI) portrays different well-known Slovak politicians as avatars/fighters in a legendary video game called Mortal Kombat. Politicians are depicted in a humorous and satirical manner. For example, you can see the portrayal of Marian Kotleba or Milan Mazurek in a uniform that resembles the Nazi uniform (which was most likely chosen due to their far-right political leaning). Even if the portrayal of some politicians might call to mind a hateful ideology, the video is purely satirical/humorous and thus, is out of scope.	No tag
	Even though the video depicts slogans/figures related to a hateful ideology, its purpose is, again, purely satirical/humorous.	No tag

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted

Harrasment and Bullying

Reminder - Public figures and Bullying

- Please be reminded that the *Bullying Statements* and *Bullying Behaviors* policies have exceptions when the target is a public figure unless the attack is a severe attack .

Exception when the target is a public figure	No exception when the target is a public figure
• Bullying Statements • Bullying Behaviors	• Revealing or Threatening to Reveal Others s Personal Data / Personally Identifiable Information • Praising or Belittling Violent Tragedies and their Victims • Threat of hacking • Sexual Harassment *(note: Exploitation & Abuse policy)*

- Politicians are viewed as public figures; therefore, bullying statements/behaviour policies should only be applied if the content contains severe attacks.
- Politicians relatives (family members) are not considered public figures (unless they acquired fame and/or notoriety for their contribution to society on their own), and all bullying policies should be applied.

1.3 Bullying Statements

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Scope	Examples
• Severe attacks on politicians ◦ Expressing hope that politicians are the victims of death, serious disease or serious injury • Bullying statement towards a family member/friend of a politician (unless they re public figures themselves) • Bullying statements toward private individuals on the grounds of their political views	• *If I could wish for one thing for Slovakia, it would be that Fico got cancer and finally dies* • *Have you seen Čaputová s daughter? She is fucking ugly!* • *Your IQ level must be under 80 if you are going to vote for Kotleba*
Another Policy	
• Other serious/credible threats - these should be controlled under "credible threats of violence" or "threats of violence"	• *I will kill Fico tomorrow*
Not in scope	Counterexamples
• Comparing a candidate to Hitler or to being a Nazi • Calling a candidate a traitor or other criminal accusations • Satire/humorous content depicting politicians in a bad light • Mocking/criticizing a candidate or current officeholder • Defamation of a politician (e.g. claim that a candidate is a drug/alcohol addict) • Politicians raising awareness of the hate they receive	

1.4 Bullying Behaviours

Scope	Examples
• Bullying behaviour toward private individuals on the grounds of their political views	• Duet depicting a user virtually punching/spitting the creator of the original video for expressing their political support for the Republika party • Showing Fico hanged

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

* Expressing hope that a politician or their family is the victim of death, serious disease, or serious injury	
Not in scope	Counterexamples
Bullying Behavior toward politicians	* Duet depicting a user virtually punching/spitting a politician s original video * Adding poop/vomit emoji to the image/video of a politician

! Within these Slovak elections, there are two viral videos (see the case bank) that show (predominantly) one-way altercations where:

* Smer politicians (inc. Kaliňák) physically attacking Matovič (OĽANO);
* Tomáš Holkovič (OĽANO) punching a man who spat onto his face.

Footage of both incidents was broadly broadcast on traditional media as well as shared on social media to document the low standards of current Slovak politics. Since such footage is of a documentary character, please approve them [no tag].

Case bank

Case	Description	Tag
Redacted	The creator wishes death to an older couple who vote for Fico	Bullying statement

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted	Smer politicians (inc. Kaliňák) physically attacking Matovič (OĽANO) – documentary content	No tag
	Tomáš Holkovič (OĽANO) punching a man who spat onto his face – documentary content	No tag
	Two high-profile politicians (Andrej Danko, Boris Kollár) verbally attacking each other. Public figures = bullying policies do not apply.	No tag

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted	Mocking two political candidates by playing chicken clucking Public figures = bullying policies do not apply	No tag
	A popular user BastaFix created a satirical song that mocks multiple Slovak politicians Public figures = bullying policies do not apply	No tag

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted

1.5 Praising or belittling violent tragedies and their victims

Scope	Examples
Note that there is no exemption for public figures under this policy title.	• User mocking politician who has been hospitalised as a result of a violent tragedy (e.g: accident, assault, rape, domestic violence, murder...)

1.6 Revealing or threatening to reveal Personal data/Personally identifiable information

Scope	Examples
Note that there is no exemption for public figures under this title, as well as under our policy title on the *threat of hacking*.	• Doxxing of politicians • Revealing home addresses of politicians and encouraging people to stalk them.
Not in scope	**Counterexamples**
Revealing common PII that is intentionally publicised by the candidates	• A candidate posting their tax returns voluntarily • Posting a campaign HQ address

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

2. Integrity and Authenticity

In the context of Slovak elections, these are two predominant misinformation trends related to:

- Russian-Ukraine War - please continue to follow *[Trending topic] Russia-Ukraine War*, which contains a classification of the most common misleading/false claims/narratives.
- George Soros - please find the classification of claims below.

Claims related to George Soros

Short profile of George Soros

- A Hungarian-born American financier and investor. Born to Jewish parents in Hungary in 1930. Soros and his family survived the Second World War by assuming a false Christian identity and changing their name from Schwartz. He later fled Communist Hungary and moved to the UK.
- Soros's subsequent success as a hedge-fund manager has made him one of the wealthiest men in the world. He is a major political donor, funding Democratic candidates and progressive causes in over 100 countries. He founded the Open Society Foundation.
- Soros's involvement in national and international politics, his wealth and his religion, have made him a target to many who disagree with him; particularly those on the populist radical-right and authoritarian leaders.

Claim	Categorisation	Additional instructions/notes/context
George Soros influences the Slovak political scene / politicians / NGOs or media.	No tag	Since Soros is a well-known political donor, his involvement in national and international politics is not misinformation, but his influence is often exaggerated.
Calling liberal politicians Sorošovci (Sorošovci = members of Soros family)	No tag	In Slovakia, liberal politicians are often labelled as Sorošovci due to fighting for progressive causes that George Soros also supports. It is considered a figure of speech.
	No tag	

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Claims that George Soros is a friend of/close to/connected to [named person/political party]		
Soros wants to bring immigrants to Europe	No tag	Since some projects backed by the Open Society Foundation (founded by Soros) support migrants, refugees and asylum seekers, it is not considered misinformation
Claims mentioning George Soros + Donation - *If true (that payment/donation was made)* * Zuzana Caputova was paid by Soros * Charta 77 received a donation from Soros	No tag	Note: As a former civil society activist, Caputová has received some Soros foundation grants in the past.
Soros is Caputova s political father Zuzana Caputova [or another high-profile Slovak politician in these elections] is constantly paid by Soros, works directly for Soros or is Soros agent	No tag General conspiracy theory	
Claims mentioning Geroge Soros + nefarious agendas; e.g. * Soros wants to bring immigrants or encourages mass immigration to destroy Europe/USA, Christian culture/values, destabilise countries, enslave white populations... * Soros is spreading LGBTQ+, gender ideology and other liberal agendas to destroy Europe/USA, Christian culture/values, destabilise countries, enslave white populations... * Soros is responsible for Covid-19 (or other viruses) * Soros destabilises states in order to provoke civil wars like Euromaidan in Ukraine or some violent protests	Dangerous conspiracy theory	
Footage/pictures depicting Soros as a global puppet master or satan	Dangerous conspiracy theory	
Soros is a former Nazi or collaborated with Nazis	Dangerous conspiracy theory	

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Case bank - George Soros

Case	Description	Tag
Redacted	Claiming that Soros collaborated with Nazis	Dangerous conspiracy theory
	Claiming that Soros is a former Nazi *The picture shows Oskar Groning and not Soros*	Dangerous conspiracy theory
	Depicting Soros as a global puppet master or satan	Dangerous conspiracy theory
	Claims mentioning Geroge Soros + nefarious agenda	Dangerous conspiracy theory

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted	Broadly connecting Soros to a certain politician or a party	No tag
	The father of Michal Šimečka (Slovak liberal politician) mentioned how Soros financial support helped liberal democracy and to beat Mečiar (a Slovak famous politician from the 90s) The video is not manipulated/fake	No tag
		No tag

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted	Criticising the financing system of NGOs while mentioning Soros	
	Claiming that Soros finances media, NGOs, elections in many countries (figure of speech: he wants to rule the world) Refusing the influence of Soros in Poland	No tag
	Claiming that Soros is a political father of Zuzana Caputova	No tag

2.1 Election Misinformation

We remove misleading content specifically related to participation in an election or other civic events.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Some election misinformation examples may include:

Scope	Examples
False or misleading information about how to vote	• You can vote from your phone • You can only vote if you are 21 years old • The day/time for voting has changed • False information about voting hours • Make sure to mark all the parties in the ballot whom you do not support • You need to provide vaccination proof to vote • You can vote with your driver s license • You need to take a photo of your vote to prove you voted • You can vote through this app • A candidate claims early victory AND encourages people to stop participating in the election
Misinformation about laws, processes, and procedures that govern the organization and implementation of elections and other civic processes	• If you vote blank, your vote will be added to the majority vote • The pen used in poll stations can be easily erased • Claims that the polls have closed early, when in reality the voting process continues.
False or misleading information about how to register to vote	• You can only register to vote if you are +25"
False or misleading information about qualifications to run in an election	• Xcandidate can t run because there is a corruption investigation against her (while it is false)
False claims about technical eligibility requirements for candidates and officials to run for office	• You need to be 35 years old to run for parliament • You need to have a PhD to be a candidate
False claims about the outcome of an election, once the results are final.	• Caputova is not the President of Slovakia, but Harabin is • The final results are not official yet

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	• We can change the final results if we sign this petition
Claims about planned election manipulation that are proved to be false	• Progresivne Slovensko (party) will manipulate the results by miscalculating the votes from Slovaks living abroad • Eset (Slovak IT company) plans to hack the elections

Not in scope

- Public interest exceptions (documentary, counter speech, satire and humorous content)
- Calls for vote Null or Blank (form of political speech)
- Calls for abstaining from voting (form of political speech)
- Video inside polling stations (even if local regulations restrict or disallow video/photography in election booths)
- Content depicting filled or unfilled ballot cards
- Content depicting a party s election programme in a misleading way
- Suggesting a candidate might win, before official results are published
- Content related to vote profiteering or vote buying
- Content related to public opinion polling

Case bank

Case	Description	Tag
Redacted	Claiming that there is a planned manipulation of elections by the Progressive Slovensko party and the well-known Slovak tech company Eset	Election misinformation
	False or misleading information about how to vote	Election misinformation

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted	Claiming that there are different voting times for different candidates (context: Czech presidential elections)	
	Even though the video names Robert Fico the winner of this election, it is only a figure of speech (showing strong support for Fico). It does not aim to reflect the actual outcome of the election.	No tag

2.2 Dangerous Misinformation

We remove false or misleading content that is likely to cause panic and could result in real world harm to a person, group, organization or institution.

Scope	Examples
Content which makes misleading claims and encourages people to destroy public property,	

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

before, during and after Election Day	* Go and destroy those ballot boxes because the Progresivne Slovensko party is manipulating elections * Steal those electronic voting machines before Eset can hack them!
Content re-purposing old videos, falsely claiming that the event or video is new in order to trigger societal panic and fear during the election	* Footage of past terrorist or violent attacks in polling stations, repurposed as current events * Footage of past criminal gang s activities in polling stations, repurposed as current events * Footage of past attacks against candidates, politicians, voters or poll workers, repurposed as current events * Footage of past military activity, repurposed as current events
False information warning of an alleged catastrophe, epidemic, or other emergency which encourages violence against individuals or groups (e.g. political opponents) during the election	* False claims of a terrorist attack at a polling station * Elections were postponed because of the earthquake. They are preparing themselves to steal the election, we need to stop this * False claims that the military is preparing to attack the population during the election * False claims regarding obstruction to electoral registration that calls for retaliatory violence * False claims based on fake documents supposedly issued by government officials warning about an imminent event that may cause panic and could result in real world harm * Content falsely claiming that a high-profile political candidate has died
Content falsely claiming basic necessities (food, water, etc) or services (banks, cash machines, etc.) are no longer available in a particular location, causing hoarding or over purchasing during the elections	* ATMs will not be available on Election Day * Public transportation will be suspended/unavailable/interrupted during Election Day
Not in scope	

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

* Public interest exceptions (documentary, counter speech) * Non-deceptive or satirical depiction of politicians look-alikes	

2.3 Dangerous Conspiracy Theory

We remove highly specific conspiracy theories that directly target a living person(s), attack a specific protected group, include a violent call to action, or deny a violent or tragic event which creates a significant amount of harm.

Scope	Examples
Content that is (i) divisive and polarizing AND (ii) dispute historical events and/or generally held belief/fact AND (iii) assign blame to a living person(s) by name.	* False theories that link candidates or political campaigns to terrorist organizations
Conspiracies that completely deny that a violent event/tragedy happened	* Conspiracies that deny abuses under dictatorships, such as assassinations and enforced disappearance * Conspiracies denying the loss of lives in violent tragedies
Any negative conspiratorial content that targets any protected groups.	* Hungarians will retaliate for the Hungarian Monarchy if Aliancia wins the parliamentary elections
Conspirational content that is accompanied by a violent call to action, and/or has previously been accompanied by a violent call to action	* They stole the election, let s close Congress

2.4 Fake News

We remove content that is packaged in a way where the user would believe the information is coming from a credible news source when it s not.

Scope	Examples	Screenshot/Video [if available]
Misleading content that is presented as news, or coming from a credible news source, when it is not		

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	* This video contains false news and is packaged to look like it is from a legitimate and trustworthy news site, when it s not	Redacted
Misleading screenshots or images of a news article that have been modified	* Front cover of The Economist is doctored, the title of the cover has been modified.	
Misleading video content with a news source banner, or with someone who appears like a professional reporter, when they are not	* This video shows someone who appears to be a professional reporter, but is not. It uses a real news outlet banner.	
Not in scope		
* Credible content by a news source * Satire, humorous content (please check if it s a parody account before actioning on the content) * News content debunking fake news * Misleading content NOT presented as news or coming from a news source		

2.5 Conspiracy Theories Related to Current, Unfolding Events

We control for theories that directly relate to a current and unfolding event where details and facts are still developing can cause moderate to significant public harm by increasing fear, panic and confusion about what is true and what is false.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Scope	Examples
Call into question the nature of present day events where details and facts are still developing and/or not confirmed.	This includes unverified electoral claims, such as: * Unverified claims that state a candidate has died suddenly or become ill * The electoral authority s website goes down on Election Day, without a verified explanation, and content provides conspirational explanations * The electoral authority takes more time than usual to provide preliminary electoral results, without a verified explanation, and content provides conspirational explanations * Unverified claims that a violent event is happening in an electoral precinct * Candidate/user/media outlet declares a candidate s early victory before an official announcement * Claims that a certain group or organization is responsible for a violent or tragic event, when the details have not been verified.
Claims of electoral fraud that cannot be verified by fact-checkers.	* Elections will be stolen or manipulated because everyone is corrupt * Voting on electronic machines can be easily altered in x location * Content that shows an individual with a badge or some other form of identification that claims to be an electoral or government official and makes claims of electoral fraud

2.6 General Conspiracy Theory

We control conspiracy theories that are not directed at a specific person(s), but target an organization or institution and create a moderate amount of harm, potentially leading to distrust in the targeted institution.

Scope	Examples
Content that calls into question a generally held belief or fact by questioning what, why or how something happened, AND suggests an institution could be responsible.	* Candidates are financed/influenced by Russia [When this is ascertained not to be the case] * The government is trying to demobilize voters by using covid lockdowns [When this is ascertained not to be the case]

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	* A fire at the Election Commission is part of a plan to alter electronic voting machines [When this is ascertained not to be the case]
Not in scope	
* Content describing unexplained events/phenomena/myths	

2.7 Impersonation (Escalate to PES/SME & CPM)

! Please immediately escalate accounts potentially impersonating politicians or political parties.

Scope	Example
* Accounts impersonating politicians/political parties/candidates (impersonating info in handle/bio/avatar) Please escalate to PES/SME & CPM	**Redacted**
Not in scope	**Counterexample**
* Partisan political accounts/fan accounts with misleading/official sounding bios * Satire/fan accounts clearly marked as such * Parody or commentary accounts clearly marked as such	

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 225

Appointment

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]
Sent: 9/25/2023 2:10:22 AM
To: CNECT-COP-TASK-FORCE@ec.europa.eu; [redacted] . [redacted]@tiktok.com]; Persa Lampropoulou (NCRTV/ERGA) [redacted]@esr.gr]; [redacted]@iag.gr; Bastien Carniel [redacted]@feedback.org]; [redacted] [redacted]@google.com]; Jelena Berkovic [redacted]@faktograf.hr]; [redacted] [redacted]@meta.com]; [redacted] [redacted]@tiktok.com]; [redacted] [redacted]@bytedance.com]; Jorge Gomes [redacted]@vosteurope.org]; Jana Kazaz [redacted]@globsec.org]; 'Tommaso Canetta' [redacted]@pagellapolitica.it]; Carnapete, Louise [redacted]@eui.eu]; Michael Meyer [redacted]@democracy-reporting.org]; Heather Dannyelle Thompson [redacted]@democracy-reporting.org]; [redacted] [redacted]@bytedance.com]; STUMP Krisztina [redacted]@ec.europa.eu]; RABBACHIN Alberto (CNECT) [redacted]@ec.europa.eu]; BRAIDA Mattia [redacted]@ec.europa.eu]; BLASCHKE Yannic [redacted]@ec.europa.eu]; GILLET Jeanne [redacted]@ec.europa.eu]
CC: KOLODZIEJSKA Anna [redacted]@ec.europa.eu]; BIRGER Albin [redacted]@ec.europa.eu]; BURIAN Peter [redacted]@ec.europa.eu]

Subject: Crisis and Elections Steering Committees Meeting
Location: Microsoft Teams Meeting

Start: 9/28/2023 6:00:00 AM
End: 9/28/2023 7:00:00 AM
Show Time As: Free

Recurrence: (none)

Produced to HJC

This Message Is From an External Sender

Dear all,

As discussed, please find below the connection details for a meeting between the steering committees of the Crisis Subgroup and the Working Group on Elections to align approaches and clarify similarities and differences between the deliverables.

Kind regards,
The CoP Team

Microsoft Teams meeting

Join on your computer, mobile app or room device
Click here to join the meeting

Meeting ID: [redacted]
Passcode: [redacted]
Download Teams | Join on the web

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Join with a video conferencing device

[redacted]@vteams.tech.ec.europa.eu
Video Conference ID: [redacted]
Alternate VTC instructions

Joining is available only from a Commission videoconferencing system.

Learn More | Help | Meeting options

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 226

Appointment

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]
Sent: 10/6/2023 7:32:34 AM
To: [redacted]

Subject: Crisis/Elections Steering Committees 2nd Meeting
Location: Microsoft Teams Meeting

Start: 10/13/2023 6:00:00 AM
End: 10/13/2023 7:00:00 AM
Show Time As: Tentative

Recurrence: (none)

Produced to HJC

This Message Is From an External Sender

Dear all,

Please find below the connection details for the second meeting of the steering committees of the Crisis Response SG and WG on Elections.

Kind regards,
The CoP Team

Microsoft Teams meeting

Join on your computer, mobile app or room device
Click here to join the meeting

Meeting ID: [redacted]
Passcode: [redacted]
Download Teams | Join on the web

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Join with a video conferencing device

[redacted]@vteams.tech.ec.europa.eu
Video Conference ID: [redacted]
Alternate VTC instructions

Joining is available only from a Commission videoconferencing system.

Learn More | Help | Meeting options

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 227

Appointment

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]
Sent: 10/13/2023 8:29:55 AM
To: CNECT-COP-TASK-FORCE@ec.europa.eu; [redacted] . [redacted]@tiktok.com]; Persa Lampropoulou (NCRTV/ERGA) [redacted]@esr.gr]; [redacted]@iag.gr; Bastien Carniel [redacted]@feedback.org]; [redacted] [redacted]@google.com]; Gori, Paula [redacted]@eui.eu]; Jelena Berkovic [redacted]@faktograf.hr]; [redacted] [redacted]@meta.com]; [redacted] [redacted]@tiktok.com]; [redacted] [redacted]@bytedance.com]; Jorge Gomes [redacted]@vosteurope.org]; Jana Kazaz [redacted]@globsec.org]; 'Tommaso Canetta' [redacted]@pagellapolitica.it]; Carnapete, Louise [redacted]@eui.eu]; Michael Meyer [redacted]@democracy-reporting.org]; Heather Dannyelle Thompson [redacted]@democracy-reporting.org]; [redacted] [redacted]@bytedance.com]; STUMP Krisztina [redacted]@ec.europa.eu]; RABBACHIN Alberto (CNECT) [redacted]@ec.europa.eu]; BRAIDA Mattia [redacted]@ec.europa.eu]; BLASCHKE Yannic [redacted]@ec.europa.eu]; GILLET Jeanne [redacted]@ec.europa.eu]; [redacted] [redacted]@microsoft.com]; [redacted] [redacted]@microsoft.com]; [redacted] (Microsoft)' [redacted]@microsoft.com]; [redacted] [redacted]@google.com]; [redacted] [redacted]@google.com]; BIRGER Albin [redacted]@ec.europa.eu]; Katarina Klingova [redacted]@globsec.org]; [redacted] [redacted]@google.com]; [redacted] [redacted]@meta.com]
CC: KOLODZIEJSKA Anna [redacted]@ec.europa.eu]; BURIAN Peter [redacted]@ec.europa.eu]; DERIVRY Tamian [redacted]@ec.europa.eu]

Subject: Crisis/Elections Steering Committees 3rd Meeting
Attachments: image001.png
Location: Microsoft Teams Meeting

Start: 10/19/2023 6:00:00 AM
End: 10/19/2023 7:00:00 AM
Show Time As: Tentative

Recurrence: (none)

Produced to HJC

This Message Is From an External Sender

Dear all,

Please find below the connection details for the second meeting of the steering committees of the Crisis Response SG and WG on Elections.

We have also uploaded the document on the Crisis Response teams channel: here

Kind regards,
The CoP Team

Microsoft Teams meeting

Join on your computer, mobile app or room device

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Click here to join the meeting

Meeting ID: [redacted]
Passcode: [redacted]
Download Teams | Join on the web

Join with a video conferencing device

[redacted]@vteams.tech.ec.europa.eu
Video Conference ID: [redacted]
Alternate VTC instructions

Joining is available only from a Commission videoconferencing system.

Learn More | Help | Meeting options

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 228

Appointment

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]
Sent: 10/23/2023 3:09:37 AM
To: CNECT-COP-TASK-FORCE@ec.europa.eu[redacted] . [redacted]@tiktok.com]; Persa Lampropoulou (NCRTV/ERGA) [redacted]@esr.gr]; [redacted]@iag.gr; Bastien Carniel [redacted]@feedback.org]; [redacted]@google.com]; Gori, Paula [redacted]@eui.eu]; Jelena Berkovic [redacted]@faktograf.hr]; [redacted]@meta.com]; [redacted]@tiktok.com]; [redacted]@bytedance.com]; Jorge Gomes [redacted]@vosteurope.org]; Jana Kazaz [redacted]@globsec.org]; 'Tommaso Canetta' [redacted]@pagellapolitica.it]; Carnapete, Louise [redacted]@eui.eu]; Michael Meyer [redacted]@democracy-reporting.org]; Heather Dannyelle Thompson [redacted]@democracy-reporting.org]; [redacted]@bytedance.com]; STUMP Krisztina [redacted]@ec.europa.eu]; RABBACHIN Alberto (CNECT) [redacted]@ec.europa.eu]; BRAIDA Mattia [redacted]@ec.europa.eu]; BLASCHKE Yannic [redacted]@ec.europa.eu]; GILLET Jeanne [redacted]@ec.europa.eu]; [redacted]@microsoft.com]; [redacted]@microsoft.com]; [redacted] (Microsoft)' [redacted]@microsoft.com]; [redacted]@google.com]; [redacted]@google.com]; BIRGER Albin [redacted]@ec.europa.eu]; 'Katarina Klingova' [redacted]@globsec.org]; [redacted]@google.com]; [redacted]@meta.com]; KOLODZIEJSKA Anna [redacted]@ec.europa.eu]; BURIAN Peter [redacted]@ec.europa.eu]; DERIVRY Tamian [redacted]@ec.europa.eu]

Subject: Crisis/Elections steering committee meeting
Location: Microsoft Teams Meeting

Start: 10/27/2023 6:00:00 AM
End: 10/27/2023 7:00:00 AM
Show Time As: Free

Recurrence: (none)

Produced to HJC

This Message Is From an External Sender

Dear all,

Please find below the connection details for the agreed meeting of the steering committees of the Crisis Response SG and WG on Elections.

Kind regards,
The CoP Team

Microsoft Teams meeting

Join on your computer, mobile app or room device
Click here to join the meeting

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Meeting ID: [redacted]
Passcode: [redacted]
Download Teams | Join on the web

Join with a video conferencing device

[redacted]@vteams.tech.ec.europa.eu
Video Conference ID: [redacted]
Alternate VTC instructions

Joining is available only from a Commission videoconferencing system.

Learn More | Help | Meeting options

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 229

Appointment

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]
Sent: 10/27/2023 7:39:29 AM
To: CNECT-COP-TASK-FORCE@ec.europa.eu; [redacted]

Subject: Crisis/Elections Steering Committees Meeting
Attachments: image001.png
Location: Microsoft Teams Meeting

Start: 11/13/2023 6:00:00 AM
End: 11/13/2023 7:00:00 AM
Show Time As: Free

Recurrence: (none)

Produced to HJC

This Message Is From an External Sender

Dear all,

As agreed, please find below the connection details for the next meeting of both committees.

We ask you to give your productive feedback on the draft Annex document by 8 November.

The agenda of the meeting:

- Discussion and agreement on the draft Annex template
- Presentation by DRI of their ideas for the "Risk assessment methodology" deliverable

Please don't hesitate to reach out if you would like to add anything,

Many thanks,

Kind regards,
The CoP Team

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Microsoft Teams meeting

Join on your computer, mobile app or room device
Click here to join the meeting

Meeting ID: [redacted]
Passcode: [redacted]
Download Teams | Join on the web

Join with a video conferencing device
[redacted]@vteams.tech.ec.europa.eu
Video Conference ID: [redacted]
Alternate VTC instructions

Joining is available only from a Commission videoconferencing system.

Learn More | Help | Meeting options

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 230

From: [Redacted] ec.europa.eu]
Sent: 11/6/2023 2:36:01 PM
To: [Redacted] [Redacted] tiktok.com]
CC: [Redacted] bytedance.com; [Redacted] ec.europa.eu]
Subject: RE: [External] NL elections meeting TikTok

Dear [Redacted]

Many thanks and looking forward to the non-paper.

From the Commission's side we will have:

- [Redacted]
-
-

From the Dutch authorities:

- [Redacted] Ministry of the Interior and Kingdom Relations
- [Redacted] Ministry of the Interior and Kingdom Relations
- [Redacted] Ministry of the Interior and Kingdom Relations
- [Redacted], Authority for Consumers and Markets

A link for the online participants will be sent in the coming days.
I'll make sure Thierry is registered. I am assuming he is aware to bring an official ID.

Kind regards,

[Redacted]

From: [Redacted] <[Redacted] tiktok.com>
Sent: Monday, November 6, 2023 3:17 PM
To: [Redacted] ec.europa.eu>
Cc: [Redacted] bytedance.com; [Redacted] ec.europa.eu>; [Redacted]; [Redacted] tiktok.com>
Subject: Re: [External] NL elections meeting TikTok

Dear [Redacted]

Many thanks for this information.

As regards attendees, the colleagues below will attend. Please note that I will be at another meeting that same morning and it is not certain that I will be out on time. However, I will do my utmost to join (virtually):

[Redacted]

May we ask who from the Commission will be attending and do you already have a list of attendees for the NL authorities?

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Also, please note that we will send you a 'non-paper' on Wednesday, outlining some of our efforts for the NL election.

Kind regards

Redacted

From: Redacted ec.europa.eu>
Date: Mon, Nov 6, 2023, 12:07 PM
Subject: RE: [External] NL elections meeting TikTok
To: Redacted tiktok.com>
Cc: "Redacted bytedance.com" Redacted bytedance.com>, Redacted ec.europa.eu>

H Redacted

They will communicate about the meeting taking place in general terms in an upcoming letter to parliament about the elections.
After the meeting they will share overarching conclusions in the letter to parliament they send to evaluate the elections.
I am assuming these conclusions will be checked with you.

Your team can contact Redacted minbzk.nl) with more questions about this.
I believe she is also in contact with Redacted about communication on a meeting which took place last week with TikTok (and others).

Kind regards,

Redacted

From: Redacted tiktok.com>
Sent: Monday, November 6, 2023 12:01 PM
To: Redacted @ec.europa.eu>
Cc: Redacted bytedance.com; Redacted ec.europa.eu> Redacted .
< Redacted tiktok.com>
Subject: Re: [External] NL elections meeting TikTok

Many thanks Redacted We will share the names later today.

Regarding this point that you flagged last week, "*the Ministry of the Interior has indicated they wish to communicate about this meeting taking place, please let us know if this raises any issues*", do you have further details on when they might communicate and what that communication will look like?

Kind regards

Redacted

From: Redacted @ec.europa.eu>

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Date: Mon, Nov 6, 2023, 11:07 AM

Subject: RE: [External] NL elections meeting TikTok

To: "Redacted tiktok.com>

Cc: "Redacted bytedance.com Redacted bytedance.com>, "Redacted ec.europa.eu>

Dear Redacted

Next to measures to mitigate the risk of mis/disinformation for these particular elections and your lessons learnt from the past elections, a specific topic of interest would also be threats against politicians, as sadly in the Netherlands this is increasingly an issue.

Could you let me know by today who will be attending in person / online from your side so I can make sure they are all registered?

A video link for any online participants should be send in the coming days.

Kind regards,

Redacted

From: Redacted Redacted tiktok.com>

Sent: Tuesday, October 31, 2023 6:07 PM

To: Redacted ec.europa.eu>

Cc: Redacted bytedance.com; Redacted ec.europa.eu Redacted Redacted tiktok.com>

Subject: Re: [External] NL elections meeting TikTok

Many thanks Redacted well noted and we will confirm all. On an initial consultation with the team, a morning slot was indeed our preference, so grateful if you could hold that slot for us until we firm up.

Are there any specific aspects that we should focus on for this meeting? We have covered a lot of the baseline work that we do in the other election meetings (SK, PL), and so any steer on particular interests on the Commission's side or indeed that of the authorities will be very helpful as we prepare.

Kind regards

Redacted

From: "Redacted ec.europa.eu>

Date: Tue, Oct 31, 2023, 6:00 PM

Subject: [External] NL elections meeting TikTok

To: "Redacted bytedance.com>, Redacted tiktok.com Redacted tiktok.com>

Cc: "Redacted ec.europa.eu>

Dear Redacted

As was already mentioned to you by Redacted we'd like to invite you to a meeting on Friday November 10th in The Hague to discuss your risk assessment and mitigation measures for the Dutch national elections of November 22nd.

The meeting will also be attended by the relevant Dutch authority (Ministry of the Interior team coordinating the approach against disinformation).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

The Ministry of the Interior has indicated they wish to communicate about this meeting taking place, please let us know if this raises any issues.

Please also let us know if you wish to also have a DSA confidential part of the meeting with only the Commission services.

The meeting will take place at the Commission's permanent representation in The Hague (Korte Vijverberg 5).

However, an option for hybrid or online participation is available.

Please let us know if a meeting at 10:00 hr works for you, and please send us the names of the attendees from your company and whether they will participate in person or online.

We encourage the participation of staff with relevant technical expertise and in-depth understanding of your enforcement processes and capacities in the area of electoral integrity to facilitate a constructive meeting.

On behalf of the DSA electoral integrity team,

Redacted

European Commission
Directorate-General for Communications Networks, Content and Technology
Unit F2: Digital Services and Platforms

Wetstraat 51
B-1049 Brussels/Belgium

Redacted.europa.eu

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 231

Non-Paper: TikTok's approach to election preparedness across the EU (with a case-study on the forthcoming Dutch snap elections)

8 November 2023

Executive Summary

- Our Global Elections Integrity programme is an overarching framework, with solutions and strategies that are developed in partnership with local experts and with the circumstances of specific elections in mind. This approach allows us to combine dedicated subject matter expert support with local teams for high-risk elections and to empower local teams by providing them with relevant tools and resources.
- We start planning months ahead of an election like the Dutch election by assembling a strong cross-functional team in order to develop robust proactive risk mitigation and response strategies based on the risk assessment and scenario planning processes.
- We enforce strict policies consisting of prohibiting election misinformation, banning political ads and restricting the activities of political parties and politicians.
- Specifically for the Netherland, our robust approach includes:
    - **Engagement with external experts.** When designing our overall programme for the Dutch election, we have considered the local risks identified by key stakeholders and experts including the Nieuwscheckers and The German Press Agency ("DPA"). We continue to actively cooperate with them on other elements including on the development of our in-app intervention.
    - **GPPPA labelling**- we classify relevant political TikTok accounts as a Government, Politician, and Political Party Account (GPPPA) and apply a number of policies that help prevent misuse of certain features such as: Incentive programs and creator monetization features, Advertising, Campaign (Election) Fundraising. We provide GPPPA account holders with information on our policies and moderation systems as well as actively encourage them to verify their profiles. We have reached out to all Dutch political parties to offer them a point of contact, if not already established, information about our security measures and verification for their accounts, whether they are active on the platform or not.
    - **Dedicated Dutch Election Hub**: We are addressing misinformation risks by providing Dutch users with official voting information though an In-App Dutch Election Hub.
    - **Media literacy campaign.** We are partnering with Nieuwscheckers to develop our media literacy strategy with the objective of educating users on how to detect misinformation and verify information they are seeing online. Nieuwscheckers educational videos will be available from our in-app Election Hub and we will redirect users to authoritative information on their website.
    - **Speaker series.** We are engaging Nieuwscheckers and DPA to share their insights and market expertise with our internal teams to inform our approach to the upcoming election.
    - **Dutch Fact-check partner**. We partner with DPA as our fact-checking expert to combat misinformation in the Netherlands.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

- **Trusted flaggers:** The ministry of the Interior and Kingdom Relations is onboarded as part of our existing trusted flagger programme. This allows them to report suspected harmful material, including election misinformation, using a dedicated email to TikTok's Risk & Response team, who review such reports. We are working hard to onboard more third party entities to this preferred reporting channel. On other topics such as discrimination or CSAM we have additional Dutch trusted flaggers in the programme.

We set out below:

- **I.** An overview of TikTok's Global Elections Integrity Programme;
- **II.** How our Elections Integrity Programme works; and
- **III.** Specific planning for the Dutch election.

I. Overview of TikTok's Global Elections Integrity Programme

Election Integrity Programme Overview: Below is a high-level overview of the Elections Integrity Programme (the "**Elections Integrity Programme**" or the "**Programme**"):

- **Context:** To deliver on its commitment to tackle Election Misinformation and to play its role in ensuring election integrity, TikTok has built an Elections Integrity Programme as part of its wider Integrity & Authenticity ("**I&A**") work. The Programme has been developed and scaled globally, and has been implemented across key elections in the EU/EEA such as the German federal election (September 2021), the French presidential election (April 2022) and legislative elections (June 2022), the Italian general election (August 2022), the Danish election (October 2022) and the recent Greek, Spanish, Slovak and Polish elections in 2023. The Programme supplements TikTok's existing Trust & Safety work, meaning that it involves multiple additional work streams and risk management practices that are implemented on a targeted basis before and during election periods. We are also publicly committed to providing transparency to users, regulators and other stakeholders around the specific measures implemented for each election in the EEA and reporting metrics around user engagement as part of our regular reporting under the **Code of Practice on Disinformation ("CoPD").**
- **Core objective and mission:** The core objective of the Programme is to ensure elections integrity on TikTok by assembling a strong cross-functional team of subject matter and local experts in order to develop robust proactive risk mitigation and response strategies. Reflecting this objective, the Programme's mission is to: (i) develop a consistent approach to election integrity across regions and cross-functional partners; (ii) empower regional teams and programme managers by providing the tools and strategies needed for local/regional success, including partnering with local / regional fact-checking partners and experts; and (iii) to foster a culture of continuous improvement across our product, process and policy verticals by identifying and addressing gaps on an ongoing basis including identifying and addressing crises relating to elections. This means that any lessons learnt from one election can be applied to further enhance our performance in subsequent elections.
- **Global framework for regional/local application:** The Programme aims to provide a global overarching framework for developing consistent and best practice approaches to

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

election integrity issues across various regions (including the EU/EEA). This includes combining dedicated subject matter expert support with local teams for high-risk elections; and empowering local teams by providing them with relevant tools and resources. The Programme is intended to provide a global framework, with solutions and strategies that can be applied and customised in partnership with local experts to the local context, nuances and circumstances of specific elections. This also means that each local team can draw on and benefit from the experience, expertise and insights drawn from elections in other regions, as part of the Programme's commitment to continuous improvement.

- **Cross-functional approach:** We very much recognise the importance of combining both subject matter expertise and local expertise and knowledge when it comes to managing misinformation risks and maintaining election integrity. The Programme entails a highly cross-functional approach, with subject matter experts across both our Trust & Safety team and multiple other teams. This involves a range of Trust & Safety functions, such as the I&A Issue Policy team, Product teams, the Partnership Outreach Management ("**OPM**") team, Risk and Response Incident Management teams, Risk Analysis teams, and Operational teams. Outside of Trust & Safety, the Programme involves global and regional Legal teams, our regional Government Relations and Public Relations teams and Communications teams, and others. There is a close working relationship with Monetisation Integrity teams.

II. How our elections integrity programme works

Upcoming preparedness: Below is a summary of the ongoing preparations for upcoming elections in the EU:

- As part of the Elections Integrity Programme, TikTok maintains an elections calendar so that relevant elections can be risk assessed well in advance of time. The goal of this exercise is to adequately: "*anticipate, prepare for, and to prevent or mitigate relevant risks*"; and involves a multi-dimensional "360" pre-election risk assessment process, incorporating input from various internal and external stakeholders. The Risk Assessment phase includes consideration of (but not limited to) the following factors: political stability, candidate safety, election significance, election interference record, relevant legal landscape and relevant market and TikTok factors.
- Having identified the relevant risks, the next step is to proceed to the scenario planning phase. Elections are unique events that may present novel policy issues and require decisive but thoughtful action from TikTok's teams under tight timelines. This phase involves assessing how the identified risks could manifest themselves on our Platform in the context of an election, across a range of areas, including at policy level, product level and processes. For example: assessing whether our existing policies cover the specific risk-related behaviours identified, and is the language in our policies sufficient to address these issues. Emerging risk issues may necessitate the need for new policy development or refinement by updating/expanding policy language to cater for additional issues/scenarios.
- Based on the risk assessment and scenario planning processes set out above, the Programme then facilitates the development of appropriate risk mitigation and response strategies. Based on the risks identified, the risk profile of the election, and other factors, the details of each strategy will vary from one election to the next.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

III. Specific planning for the Dutch election

Early general elections are scheduled to be held in the Netherlands on 22 November 2023 to elect the members of the House of Representatives. The election had been expected to be held in 2025, but a snap election was called after the cabinet of MP Rutte collapsed on 7 July 2023.

In the specific context of the Dutch election, this approach involves the following key timelines:

- **7 July 2023:** The snap election was called after cabinet collapsed
- **19 July:** Cross-functional (XFN) Dutch Election Working Group convened for detailed planning purposes.
- **July-August-September-October 2023:** Ongoing intensive election preparations, scenario-planning and engagement, as further described below.
- **25 October**: Launch of our in-app Dutch Election Hub
- **6 November:** Netherlands election blog posted on TikTok Newsroom
- **22 November 2023:** Scheduled election day for Dutch general election.
- **Post-election:** Ongoing monitoring for post-election risks and election debrief.

Our preparations for the Dutch election are ongoing and, as part of a multi-prong approach, we currently plan to implement the following measures:

Policy measures:

- **Community Guidelines:** TikTok's Community Guidelines are the starting point when it comes to how we frame and shape our content moderation strategies in relation to election misinformation. Under the heading of "*Civic and Election Integrity*", TikTok's Community Guidelines makes clear that we do not allow misinformation about civic and electoral processes, regardless of intent. This includes misinformation about how to vote, registering to vote (where applicable), eligibility requirements of candidates, the processes to count ballots and certify elections, and the final outcome of an election. We remove this content from our platform and for unverified claims about the outcome of an election that is still unfolding and may be false or misleading, we make that content ineligible for the For You feed. In our six monthly COPD report, we report on the number of videos we remove that are deemed harmful misinformation.

- **Covert Influence Operations:** TikTok's integrity and authenticity policies do not allow coordinated attempts to influence or sway public opinion while also misleading individuals, our community, or our systems about an account's identity, approximate location, relationships, popularity, or purpose. When we investigate and remove these operations, we focus on behaviour and assessing linkages between accounts and techniques to determine if actors are engaging in a coordinated effort to mislead TikTok's systems or our community. In each case, we believe that the people behind these activities coordinate with one another to misrepresent who they are and what they are doing.

- **Ban on political ads:** As stated in the "*Civic and Election Integrity*" section of the Community Guidelines, TikTok does not allow political ads on the platform. All ads are reviewed before being displayed on our platform. Upon review, if an ad is deemed to violate our strict ads policies including our prohibition of political ads, it will not be permitted on our platform. We also review ads reported to us and upon review, any violating ads will be

removed. In our six monthly COPD reports, we have committed to further increasing transparency around our approach to political ads and to report on the number of removals of political ads and appeals.

- **Restrictions for Government, Politician, and Political Party Accounts (GPPPAs):** As explained in TikTok's political ads policy,[1] TikTok's priority in this space is to keep harmful election misinformation off the Platform and ensure our community has a positive and welcoming experience. That's why we classify relevant political TikTok accounts as a Government, Politician, and Political Party Account and apply a number of policies that help prevent misuse of certain features. We have also reached out to Dutch political parties to offer resources on our security measures and to offer support in the form of verification as well as to provide a single point of contact in case of queries or incidents regarding e.g. content, online threats or hate speech directed at their accounts.

  For designated GPPPAs, the following TikTok features are not available:

  - **Prohibition on political advertising:** TikTok has long prohibited political advertising, including both paid ads and creators being paid to make branded political content. This means that accounts we identify as belonging to politicians and political parties have their access to advertising features turned off.

  - **Campaign (Election) Fundraising:** Just as political advertising doesn't fit with our goal of making TikTok a place that brings people together, neither does campaign fundraising. Solicitations for campaign fundraising by GPPPAs are not permitted on the platform. That includes content like a video from a politician asking for donations, or a political party directing people to a donation page on their website.

  We use the same content moderation practices on GPPPAs as we do for other TikTok accounts. This means we'll remove any violative content and permanently remove the account for any single severe content violation, such as showing real-world violence or torture. However, because of the role these public interest accounts play in civic processes and civil society, we enforce different account restrictions in keeping with our commitment to human rights and free expression. Further information is available here about how we enforce restrictions on public interest accounts.

Stakeholder engagement and partnerships:

As part of our Election Integrity Programme, our teams undertake extensive engagement with external stakeholders, including with local subject matter experts, electoral commissions, and suitably qualified partners, as a means of risk identification and to further understand and monitor election-related risks. In the context of the Dutch election, this has involved:

- **Fact-checkers:** We have fact-checking coverage in 17 official European languages including Dutch and, therefore, the spoken language of 23 EEA markets. We are working together with our IFCN certified fact-checking partner, DPA, to help support our election integrity efforts in The Netherlands and tackle any election misinformation on the platform. We have recently increased our ability to receive pro-active flagging of potential misinformation from our fact-checking partners. Our partners are authorised to proactively identify content that may constitute harmful misinformation on our platform and suggest prominent misinformation that is circulating on other social media platforms or websites that may benefit from verification. Our

[1] **Note:** Updating our policies for political accounts, Newsroom, October 2022.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

fact-checking partners can also provide us with regular reports identifying general misinformation trends observed on our platform and across the industry generally, including new/changing industry or market trends, events or topics that generated particular misinformation or disinformation.

- **Trusted flaggers:** TikTok has developed meaningful collaborations with third-party entities (e.g. NGOs and regulators) as part of our existing trust . These qualified collaborations entail the creation of a preferred reporting channel dedicated to such partners, who can report suspected harmful material using a dedicated email to TikTok's Risk & Response team, who review such reports. In the Netherlands the ministry of the Interior and Kingdom Relations is a trusted flagger. Next to trusted flaggers, we have a close working relationship with the Dutch Police through their SPOC centre in Driebergen for both standard investigations as well as through an emergency channel for threat cases aimed at people or places. We also share cases proactively through this channel.

- **Speaker Series:** As part of the Election Integrity Programme, TikTok's OPM teams organises an external Election Speaker Series which involves election-related speakers in the lead-up to elections. As part of this, we invite suitably qualified external local/regional experts to share their insights and market expertise with our internal teams in order to inform our approach to the upcoming election. In preparation for the Dutch election, TikTok engaged DPA and Nieuwscheckers to participate in the Speaker Series.

Product measures:

Together with systematically removing content from our platform that violates our I&A policies, we dedicate significant resources to utilising in-app intervention tools to show users additional context on certain content or redirect them to authoritative information. These tools include video notice tags and search interventions, which are made available in 23 EU official languages including Dutch.

- **Dutch Election Hub:** We will address misinformation risk by providing Dutch users with official voting information though an In-App Dutch Election Hub (voter information centre):
  - **Reliable information:** The Election Hub will contain information about the upcoming election sourced from official Dutch governmental web-resources.
  - **Key messages:** Our main goal and focus is to inform Dutch users about the election process, including but not limited to:
    - When to vote
    - Where to vote
    - Process of voting
  - **Media literacy:** In addition to providing information from reliable sources, we are partnering with digital literacy partner Nieuwscheckers to educate Dutch users on how to detect misinformation and verify information they are seeing online. The educational videos are available to Dutch users from the Election Hub and at the TikTok account of Nieuwscheckers.
  - **Access to Election Hub:** We continuously label election-related videos enabling users to directly access the Election Hub, these labels are visible in the For-You Feed. Labels are placed based on matching election-related keywords in the video's title. In addition, users

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

searching for election-related content on TikTok will be served an election banner enabling them to directly access the Election Hub.

- "**Unverified content label**": As well as the tools we deploy to increase users' digital citizenship regarding specific topics, we add unverified content labels to content related to unfolding or emergency events that may arise including in the context of elections, including content which has been assessed by our fact-checkers but cannot be verified as accurate (see screenshots at the **Annex**). The purpose of these labels is to raise users' awareness about the credibility of content and to encourage people to reconsider sharing such content. When an unverified content label is applied:

    - The video will become ineligible for recommendation into anyone's For You feed to limit the spread of potentially misleading information.

    - The video's creator is also notified that their video was flagged as unsubstantiated content and is provided additional information about why the warning label has been added to their content. This is again to raise the creator's awareness about the credibility of the content that they have shared. We disclose metrics around user engagement with these tools in our six monthly COPD reports.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Annex

Dutch in-app Election Hub

We launched a search intervention and in-app election hub to provide users with up to date information in advance of the Dutch election.

Election Hub

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

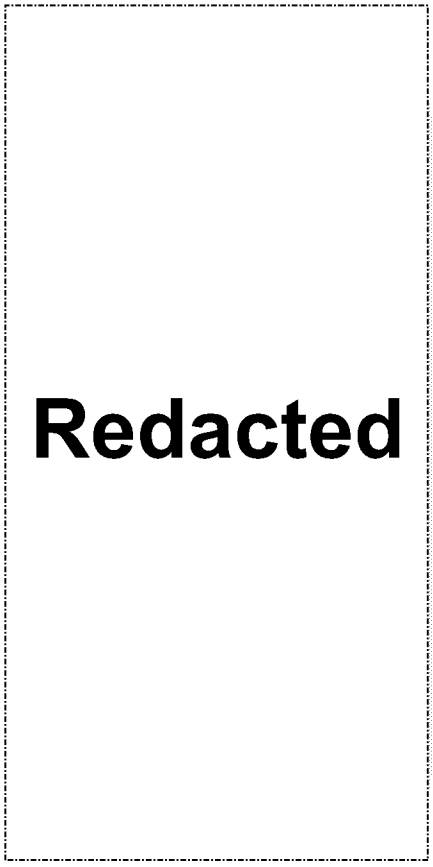
TT_HJC_006259

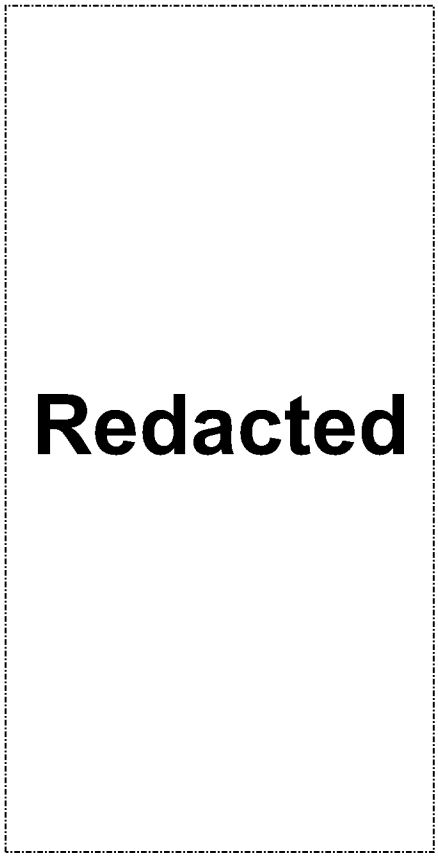
Redacted

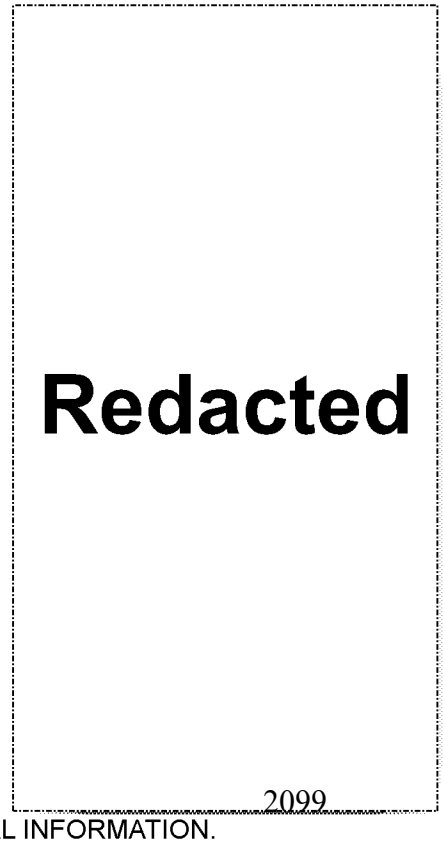
CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Unverified content labels

To raise users' awareness about the credibility of content and to encourage people to reconsider sharing such content, we add warning labels to content related to unfolding or emergency events, including content which has been assessed by our fact-checkers but cannot be verified as accurate







CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 232

Appointment

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]
Sent: 11/14/2023 5:49:01 AM
To: CNECT-COP-TASK-FORCE@ec.europa.eu; [redacted]@tiktok.com]; Persa Lampropoulou (NCRTV/ERGA) [redacted]@esr.gr]; [redacted]@iag.gr; Bastien Carniel [redacted]@feedback.org]; [redacted]@google.com]; Gori, Paula [redacted]@eui.eu]; Jelena Berkovic [redacted]@faktograf.hr]; [redacted]@meta.com]; [redacted]@tiktok.com]; [redacted]@bytedance.com]; Jorge Gomes [redacted]@vosteurope.org]; Jana Kazaz [redacted]@globsec.org]; 'Tommaso Canetta' [redacted]@pagellapolitica.it]; Carnapete, Louise [redacted]@eui.eu]; Michael Meyer [redacted]@democracy-reporting.org]; Heather Dannyelle Thompson [redacted]@democracy-reporting.org]; [redacted]@bytedance.com]; STUMP Krisztina [redacted]@ec.europa.eu]; RABBACHIN Alberto (CNECT) [redacted]@ec.europa.eu]; BRAIDA Mattia [redacted]@ec.europa.eu]; BLASCHKE Yannic [redacted]@ec.europa.eu]; GILLET Jeanne [redacted]@ec.europa.eu]; [redacted]@microsoft.com]; [redacted]@microsoft.com]; [redacted] (Microsoft)' [redacted]@microsoft.com]; [redacted]@google.com]; [redacted]@google.com]; BIRGER Albin [redacted]@ec.europa.eu]; 'Katarina Klingova' [redacted]@globsec.org]; [redacted]@google.com]; [redacted]@meta.com]; KOLODZIEJSKA Anna [redacted]@ec.europa.eu]; BURIAN Peter [redacted]@ec.europa.eu]; DERIVRY Tamian [redacted]@ec.europa.eu]

Subject: Crisis-Elections Steering Committees Meeting
Attachments: image001.png
Location: Microsoft Teams Meeting

Start: 11/20/2023 6:00:00 AM
End: 11/20/2023 7:00:00 AM
Show Time As: Tentative

Recurrence: (none)

This Message Is From an External Sender

Dear all,

Please find below the connection details for the next meeting of the Crisis-Elections steering committees and, as agreed yesterday, a quick recap' of the next steps:

- Outstanding points on the Annex to the Rapid Response System will be discussed offline by the EC-Clement-Lara-Michael
- Next meeting of the Crisis-Elections steering committees on 20/11. Agenda:
  - Finalisation of the Annex to the RRS (here)
  - Rapid Response System (Google's proposal: here)

If this meeting is successful in agreeing on the common approach, we aim to divide the two steering committees back (one on Crisis and one on Elections) for future meetings.

The work on the deliverables (Rapid Response System and Risk Assessment) will then continue by the two respective groups to cater to the specificities of these situations.

Produced to HJC

Let us know if you have any question and thank you again for your engagement in this important work strand,

Kind regards,
The CoP team

Microsoft Teams meeting

Join on your computer, mobile app or room device
Click here to join the meeting

Meeting ID: [redacted]
Passcode: [redacted]
Download Teams | Join on the web

Join with a video conferencing device
[redacted]@vteams.tech.ec.europa.eu
Video Conference ID: [redacted]
Alternate VTC instructions

Joining is available only from a Commission videoconferencing system.

Learn More | Help | Meeting options

Produced to HJC

Exhibit 233

Appointment

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]
Sent: 11/30/2023 4:19:19 AM
To: CNECT-COP-TASK-FORCE@ec.europa.eu; BIRGER Albin [redacted]@ec.europa.eu]; BLASCHKE Yannic [redacted]@ec.europa.eu]; BRAIDA Mattia [redacted]@ec.europa.eu]; BURIAN Peter [redacted]@ec.europa.eu]; Carlos Hernandez [redacted]@maldita.es]; [redacted]@tiktok.com]; [redacted]@microsoft.com]; [redacted]@bytedance.com]; [redacted]@google.com]; Dominika Hadju [redacted]@globsec.org]; FILIPEK Jon [redacted]K@ec.europa.eu]; [redacted]@bytedance.com]; GILLET Jeanne [redacted]@ec.europa.eu]; [redacted]@linkedin.com]; Heather Dannyelle Thompson [redacted]@democracy-reporting.org]; Ines Narciso [redacted]@vosteurope.org]; Ines Narciso 2 [redacted]@pm.gov.pt]; Irene Larraz [redacted]@newtral.es]; Jana Kazaz [redacted]@globsec.org]; Jelena Berkovic [redacted]@faktograf.hr]; Joao Ribeiro [redacted]@vosteurope.org]; Jorge Gomes [redacted]@vosteurope.org]; Julian Merer [redacted]@arcom.fr]; Katarina Klingova [redacted]@globsec.org]; Katya Viadziorchyk [redacted]@aiforensics.org]; KOLODZIEJSKA Anna [redacted]@ec.europa.eu]; [redacted]@meta.com]; [redacted]@logically.ai]; [redacted]@bytedance.com]; [redacted]@google.com]; [redacted]@google.com]; Marc Faddoul [redacted]@aiforensics.org]; [redacted]@meta.com]; [redacted]@twitch.tv]; [redacted]@microsoft.com]; [redacted]@linkedin.com]; Michael Meyer [redacted]@democracy-reporting.org]; [redacted]@google.com]; Nicolas Faucouit [redacted]@arcom.fr]; Omri Preiss [redacted]@alliance4europe.eu]; Pawel Terpilowski [redacted]demagog.org.pl]; RABBACHIN Alberto (CNECT) [redacted]@ec.europa.eu]; Salvatore Romano [redacted]@aiforensics.org]; Sam Jeffers [redacted]@whotargets.me]; [redacted] (google) [redacted]@google.com]; [redacted]@bytedance.com]; STUMP Krisztina [redacted]@ec.europa.eu]; [redacted]@eeas.europa.eu; [redacted]@microsoft.com]; Tommaso Canetta [redacted]@pagellapolitica.it]; [redacted]@bytedance.com]; Víctor Rico [redacted]@maldita.es]

Subject: WG on Elections Meeting
Attachments: image001.png
Location: Microsoft Teams Meeting

Start: 12/5/2023 7:00:00 AM
End: 12/5/2023 8:00:00 AM
Show Time As: Free

Recurrence: (none)

Produced

This Message Is From an External Sender

Dear Members of the WG on Elections,

We would like to call a meeting on Tuesday 5 December 16:00-17:00. Please find the connection details below.

The agenda will be two-fold:

- Presenting the work carried out over the last weeks by the steering committee: a document which will serve both as a reporting template for elections matters and as an annex of the rapid response system. You can find a clean copy on the WG's Teams space here.
- Starting the conversation on the Risk Assessment Methodology

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

We are looking forward to discussing with you these two important deliverables,

Kind regards,
The CoP Team

Microsoft Teams meeting

Join on your computer, mobile app or room device
Click here to join the meeting

Meeting ID: [redacted]
Passcode: [redacted]
Download Teams | Join on the web

Join with a video conferencing device
[redacted]@vteams.tech.ec.europa.eu
Video Conference ID: [redacted]
Alternate VTC instructions

Joining is available only from a Commission videoconferencing system.

Learn More | Help | Meeting options

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 234

Appointment

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]
Sent: 11/30/2023 3:39:32 AM
To: [redacted]@tiktok.com]; Persa Lampropoulou (NCRTV/ERGA) [[redacted]@esr.gr]; [redacted]@iag.gr; Bastien Carniel [[redacted]@feedback.org]; [redacted] [[redacted]@google.com]; Gori, Paula [[redacted]@eui.eu]; Jelena Berkovic [[redacted]@faktograf.hr]; [redacted]@meta.com]; [redacted] [[redacted]@tiktok.com]; [redacted]@bytedance.com]; Jorge Gomes [[redacted]@vosteurope.org]; Jana Kazaz [[redacted]@globsec.org]; 'Tommaso Canetta' [[redacted]@pagellapolitica.it]; Carnapete, Louise [[redacted]@eui.eu]; Michael Meyer [[redacted]@democracy-reporting.org]; Heather Dannyelle Thompson [[redacted]@democracy-reporting.org]; [redacted] [[redacted]@bytedance.com]; STUMP Krisztina [[redacted]@ec.europa.eu]; RABBACHIN Alberto (CNECT) [[redacted]@ec.europa.eu]; BRAIDA Mattia [[redacted]@ec.europa.eu]; BLASCHKE Yannic [[redacted]@ec.europa.eu]; GILLET Jeanne [[redacted]@ec.europa.eu]; [redacted] [redacted]@microsoft.com]; [redacted] [[redacted]@microsoft.com]; [redacted] (Microsoft)' [[redacted]@microsoft.com]; [redacted]@google.com]; [redacted] [redacted]@google.com]; BIRGER Albin [[redacted]@ec.europa.eu]; 'Katarina Klingova' [[redacted]@globsec.org]; [redacted]@google.com]; [redacted] [redacted]@meta.com]; KOLODZIEJSKA Anna [[redacted]@ec.europa.eu]; BURIAN Peter [[redacted]@ec.europa.eu]; DERIVRY Tamian [[redacted]@ec.europa.eu]

Subject: Crisis-Elections Steering Committees Meeting
Attachments: image001.png
Location: Microsoft Teams Meeting

Start: 12/14/2023 7:00:00 AM
End: 12/14/2023 8:30:00 AM
Show Time As: Tentative

Recurrence: (none)

Produced to HJC

This Message Is From an External Sender

Dear Members of the joint Steering Committees,

As discussed in our last meeting, we are sending you a save the date for the next joint meeting, during which we will discuss the rapid response system on the basis of the draft prepared by Google (c.f. here).

In the meantime, as agreed, the Crisis SG and WG on Elections will validate the Annex/reporting template (c.f. here) for their respective purposes (crises and elections) in individual separate meetings, for which we will send invitations shortly.

Kind regards,

The CoP Disinfo Team

Microsoft Teams meeting

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Join on your computer, mobile app or room device
Click here to join the meeting

Meeting ID: [redacted]
Passcode: [redacted]
Download Teams | Join on the web

Join with a video conferencing device
[redacted]@vteams.tech.ec.europa.eu
Video Conference ID: [redacted]
Alternate VTC instructions

Joining is available only from a Commission videoconferencing system.

Learn More | Help | Meeting options

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 235

Appointment

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]
Sent: 12/14/2023 8:10:41 AM
To: [redacted]@meta.com]

Subject: FW: WG on Elections - Meeting on Risk Assessment Methodology
Location: Microsoft Teams Meeting

Start: 12/20/2023 6:00:00 AM
End: 12/20/2023 7:00:00 AM
Show Time As: Tentative

Recurrence: (none)

This Message Is From an External Sender

Produced to HJC

-----Original Appointment-----
From: CNECT COP TASK FORCE <CNECT-COP-TASK-FORCE@ec.europa.eu>
Sent: Friday, December 8, 2023 11:10 AM
To: CNECT COP TASK FORCE; [redacted]; 'Tommaso Canetta - PagellaPolitica'; STUMP Krisztina (CNECT); RABBACHIN Alberto (CNECT); BIRGER Albin (CNECT); BLASCHKE Yannic (CNECT); BRAIDA Mattia (CNECT); GILLET Jeanne (CNECT); DERIVRY Tamian (CNECT); [redacted]
Cc: KOLODZIEJSKA Anna (CNECT); BURIAN Peter (CNECT)
Subject: WG on Elections - Meeting on Risk Assessment Methodology
When: mercredi 20 décembre 2023 15:00-16:00 (UTC+01:00) Brussels, Copenhagen, Madrid, Paris.
Where: Microsoft Teams Meeting

Dear all,

Thank you for having expressed interest in taking the work of the WG on Elections forward- in particular by advancing on the risk assessment methodology.

We propose a first meeting to kick start the discussion on 20 December 2023, 3-4pm CET, please find the connection details below.

Kind regards,
The CoP team

Microsoft Teams meeting

Join on your computer, mobile app or room device
Click here to join the meeting

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Meeting ID: [redacted]
Passcode: [redacted]
Download Teams | Join on the web

Join with a video conferencing device

[redacted]@vteams.tech.ec.europa.eu
Video Conference ID: [redacted]
Alternate VTC instructions

Joining is available only from a Commission videoconferencing system.

Learn More | Help | Meeting options

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 236

ROMANIAN INFORMATION SERVICE

NOTE NO. 2 TO THE ROMANIAN SUPREME COUNCIL FOR NATIONAL DEFENSE

I. **Actions of a cyber state actor** on the IT&C infrastructures supporting the electoral process, hosted by the Permanent Electoral Authority (AEP) and the Special Telecommunications Service (STS).

By using specific methods, on 24.11.2024, the Romanian Intelligence Service (SRI) obtained data regarding the publication of access credentials associated with "**bec.ro**" (i.e. Romanian Central Electoral Office - BEC - website), "**roaep.ro**" (i.e. AEP website) and "**registrulelectoral.ro**" (i.e. Romanian Electoral Registry/AEP website) within several Russian origin cybercrime platforms, similar data being also identified within a private Telegram channel well known for disseminating exfiltrated data from many states, except the Russian Federation.

Following the investigations initiated, it was determined that the exfiltration was carried out either by targeting legitimate users to whom the user/password credentials were distributed, or by exploiting the legitimate training server made available by STS at **https://operatorsectie.roaep.ro**.

Regarding the infrastructure topology, STS manages the main sequence related to the voting process: recording voter turnout, ensuring the accuracy of ballot counting by video recording the process of opening the ballot boxes and counting the votes, and centralizing the results.

The infrastructure sequence managed by AEP serves the real-time display of voter turnout, statistics on vote distribution by various criteria (categories, age, gender, urban/rural environment, etc.), as well as making the electoral legislation available.

These posts were published after a cyber incident targeted and affected the AEP IT&C infrastructure on 19.11.2024, through which cyber attackers managed to compromise a map server (***gis.registrulelectoral.ro***), connected both externally, to the internet, and to AEP's internal network.

In this context, **a large number of cyber-attacks**[1] **(over 85,000)** were identified, which aimed to exploit existing vulnerabilities in the IT systems supporting the electoral process, to **gain access to data in the IT systems, alter their integrity, change the content presented to the public and make the infrastructure unavailable.**

The National Cyberint Center conducted technical assessments on related information systems by analysing the log files for the period 20-26.11.2024, generated by the cybersecurity equipment used by:

- *prezenta.roaep.ro* - platform for monitoring and displaying statistics on voter turnout;
- *voting_roaep.ro* - blockchain transaction platform;
- *prezidentiale1-sicpv.bec.ro* - computer system for centralizing minutes;
- *simpv.bec.ro* - computer system for monitoring voter turnout;
- *simpv.roaep.ro* - computer system for monitoring voter turnout;
- *simpv.stsnet.ro* - computer system for monitoring voter turnout.

[1] 1 Type:
- SQL injection (SQLi) - Attack that involves injecting SQL-type malicious code into an application to access and/or modify the database behind it;
- Cross Site Scripting (XSS) - Attack that exploits a vulnerability found in a web page and allows an attacker to insert lines of code into web pages visited by other users (victims), to obtain data with restricted access.

The attacks in question continued in a **sustained manner, including on election day and the night after the election (25.11.2024). Computer systems from over 33 countries** were used to launch the attacks, using advanced anonymization methods to make the attribution process difficult.

* It has to be underlined that specific investigations have been launched together with the AEP and STS. As the assessment of the cyberattack is ongoing, we currently do not have any clear data on the attacker or on the impact on the electoral process.

The modus operandi, as well as the scale of the cyber campaign, lead to the conclusion that the attacker has considerable resources that are **correlated with specific methods used by a state-sponsored attacker**. At the same time, the AEP infrastructure still remains affected by vulnerabilities which, to the extent that they are exploited by attackers, can escalate access within the network and ensure persistence.

II. In the context of the issues circulating in the online environment, the obtained data revealed that the reason for the massive and accelerated growth of **Călin GEORGESCU**'s popularity on the TikTok social platform is due to a very well-organized promotional campaign.

Călin GEORGESCU benefited from preferential treatment on the TikTok platform, because the content he posted **was not marked as belonging to a candidate**, which favoured mass dissemination, as the published videos were not officially associated with the electoral campaign.

Consequently, **his visibility increased preferentially compared to the other candidates, whose posts were massively filtered, exponentially diminishing their online presence.**

This preferential treatment was exacerbated by **TikTok's failure to comply with BEC Decision** no. 175D of 20.11.2024, which, in art. 3, ordered "the removal of electoral propaganda materials from the online environment that illustrate the candidate Călin Georgescu in the 2024 Romanian Presidential elections, which do not contain the identification code of the fiscal representative".

The request was sent to TikTok, through AEP, on 21.11.2024, 08:00 hrs. Subsequently, at TikTok's request, a return was made with the CMF code, which following the analysis made by AEP was not found in any of the candidate's posts.

On 22.11.2024, 13:47 hrs, TikTok sent AEP confirmation of the removal of the posts that are the subject of BEC Decision no. 175D of 20.11.2024, by blocking visual access to them from the territory of Romania, **they still remaining visible in other states and being possible to be distributed.**

Meanwhile, according to information obtained from TikTok representatives, an analysis was carried out regarding the online activities subsumed under the Călin GEORGESCU promotion campaign.

TikTok's first notification of the fact that a Călin GEORGESCU promotion campaign was underway took place in 2020, and in 2021, it was reported by them (most likely to TikTok management) as a suspicious activity.

The conclusions of the current analysis reveal the following:

• Telegram and Discord channels were discovered where they discussed how to coordinate and

avoid being blocked on the platform, so no direct link was identified between the multiple TikTok accounts used to promote Călin GEORGESCU, given that the activity was carried out from multiple geolocations;

• **the activity of the accounts was allegedly coordinated by a state actor**, who would have used an alternative communication channel to "roll" messages on the platform;

• does not use bot farms on the platform, but operates more discreetly from the outside, so as not to violate the platform's usage policies;

• those involved in the campaign to promote the person in question demonstrate a very good knowledge of TikTok's security policies, and have the necessary know-how to circumvent them;

• there is a very good digital marketing company behind it;

• the social accounts that promoted Călin GEORGESCU on TikTok disseminated identical messages, without any coordination on the platform (no fingerprints were detected that would connect the devices used);

The growth of accounts was not organic (similar to natural viral events), so TikTok believes that they are basically coordinated volunteers ("**mass guerrilla political campaign**" or "**brute force attack in cybersecurity**").

• the dissemination of messages within the TikTok platform was carried out in swarms (swarming).

Also, during the last days, TikTok has identified a massive promotional activity, carried out in the last two weeks, in support of POT (Young People's Party), a sovereigntist party, founded in 2023, which supports Călin GEORGESCU.

Exhibit 237

ROMANIAN INFORMATION SERVICE

NOTE NO. 1 TO THE ROMANIAN SUPREME COUNCIL FOR NATIONAL DEFENSE

In the context of the first round of the presidential elections, data were obtained that revealed an aggressive promotional campaign, carried out in circumvention of national electoral legislation, as well as the exploitation of algorithms of social media platforms for **the accelerated growth** of Călin Georgescu.

I. Using TikTok to increase Călin Georgescu's notoriety

His rise in the opinion polls was determined by a **coordinated campaign** to increase his popularity, **amplified 2 weeks before** the polling date, in particular **on the TikTok platform**, which managed to ensure his victory in the first round with 22.94%.

Călin GEORGESCU was at a very low level of notoriety a few weeks before the elections. In polls on voting intention between October 30 and November 5, Călin GEORGESCU was quoted at less than 1%. In polls from **November 15-20**, he was **ranked 6th, with 6.2%** voting intention, and on **November 20-21**, polls put Călin GEORGESCU at **10.6%.**

The **massive promotion activity** involved a campaign on **TikTok**, through several coordinated accounts that actively published electoral content, both with the help of recommendation algorithms and paid promotion.

1.1 The network of accounts directly associated directly with Călin GEORGESCU's campaign initially consisted of **25,000** accounts on the TikTok platform, which became **very active two weeks before the election date**. The promotional campaign was **particularly well organized**, with the number of followers increasing significantly.

It was noted that **797** of the accounts that initially formed the support network had been created **since 2016**. They had very low activity (1%) until 11.11.2024, when the entire network was fully activated.

Of particular importance is that no shared technical resources were used and no shared IP addresses (IP sharing) were observed. In concrete terms, the TikTok accounts in question were assigned unique IP addresses, which denotes a modus operandi designed to facilitate the identification of the extent of the network based on the use of a common infrastructure (characteristic of bots).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

1.2 The activity of the accounts in the network was organized Outside the TikTok platform, namely on the Telegram instant messaging application.

The Telegram channel *"Propagator- implică-te și tu" (Propagator - get involved), "Renașterea României" (Rebirth of Romania)," Hrană, Apă, Energie" (Food, Water, Energy)* Redacted was identified, which has the role of **coordinating other users on the posts and video content distributed.**

The Redacted subscribers received clear instructions on how to act on the accounts, recommendations on user behavior at the TikTok level, the posts that would be promoted or distributed, in terms of including specially chosen *emoji* and the name of the candidate to exploit TikTok algorithms. These were prepared in advance and instantly uploaded to TikTok.

> Thus, posts encouraging followers to post certain pre-defined tags (,,#CG", ,,#diaspora", ,,#calingeorgescu"), to distribute certain videos posted on the groups, and to distribute certain videos posted on the groups, were identified on the Telegram groups, as well as indications to circumvent the verification mechanisms of the TikTok platform (screen recording and altering content to be perceived by the platform as original content).

On 24.11.2024, @propagatorcg had **1,088 subscribers**, and on 01.12.2024 it had 5,005 **users**, an increase of 603 members from the previous day.

According to the data provided by TikTok on 28.11.2024, in dialog with the Romanian authorities, after the creation of the Telegram channel Redacted 76 **new TikTok accounts were registered, having the exclusive role of promoting candidate Călin GEORGESCU.**

The TikTok representatives also stated that Redacted **was created on June 15, 2024, and had, as of 28.11.2024, 3,755 subscribers who were trained on distributing videos and posts** and leaving comments, using certain hashtags, on Facebook, YouTube and TikTok platforms.Thereby, two patterns were detected regarding the candidate's promotion activity:

1.1183 accounts used,, #CG11" in all videos posted on TikTok. Out of these, 157 were registered on TikTok after the creation of the Telegram channel, and 76 were created and used exclusively with the intention of posting promotional content, the rest were associated with real people who voluntarily posted pro-Călin GEORGESCU content.

2.The accounts used a string of specific emoticons accompanied by the candidate's name to exploit the platform's algorithms. TIK TOK representatives identified 83 accounts that followed this pattern, which left 2912 comments (23% of which were posted by the official account of Călin GEORGESCU).

TikTok also reported that it had identified a potential network of accounts involved in manipulating public opinion in an electoral context, as well as a network of accounts affiliated with the Russian publication **Sputnik** that targeted users in Romania and Moldova.

It is interesting to note that some of the Telegram groups in which Călin GEORGESCU's supporters have coagulated **were created in September 2022.** As an example, the group *Călin Georgescu-Proiect de țară* (Călin Georgescu - Country Project) was created, where the candidate's narrative was promoted, together with four other similar Telegram groups.

1.3 In promoting Călin GEORGESCU, an extensive network of high-profile public figures on the TikTok platform (influencers) was used to promote his campaign both directly (by publicly supporting the candidate) and indirectly (through messages that appeared neutral but contained tags associated with the candidate, such as *#echilbrușiverticalitate*). Some of them intentionally did not mark the advertising as paid, in order to avoid associating the posts with election campaigning.

A non-exhaustive list of public persons who expressed their support for Călin GEORGESCU reveals the following accounts at the TikTok level: Redacted - Redacted Redacted

1.4 Creation and use of accounts falsely representing Romanian state institutions in support of Călin Georgescu

Subsequently, on the margins of the parliamentary elections of December 1, **dozens of TikTok accounts were found falsely using the SRI logo and the Anti-Terrorist Brigade (BAT),** respectively displaying thousands of followers and over 100,000 likes.

On these accounts, **numerous images** of Romanian Police forces and BAT, taken from the public space, **were distributed.** In the comments, **appreciative posts were identified on the SRI-BAT/BAT/law enforcement institutions and in support of candidate GEORGESCU, thus falsely inducing the idea that state institutions support him.**

1.5 TikTok's conduct and relations with the Romanian authorities

On November 20, 2024, by decision no.175D, the Central Electoral Bureau (BEC) ordered *"the removal of electoral propaganda materials from the online environment depicting the candidate Călin Georgescu in the 2024 elections for the President of Romania, which do not contain the identification code of the fiscal trustee".*

The request was sent to TikTok, via the Permanent Electoral Authority (AEP), on 21.11.2024 (08:00).

On 22.11.2024 (at 13:47), **TikTok sent AEP confirmation of the removal of the postings subject** to BEC Decision no. 175D of 20.11.2024, **by blocking visual access to them from Romania,** they remain visible in other countries and can be distributed.

However, subsequent checks revealed that **TikTok did not delete the electoral content as requested by AEP, and it continued to be available to the Romanian public even after the end of the electoral campaign, including on election day** (in contravention of Romanian law).

Posts with purely electoral content from accounts associated with the network promoting Călin Georgescu were categorized by TikTok as **entertainment posts** and were displayed:

- unrestricted to a wide range of users, although this is **contrary to the Chinese company's own policies on political content;**

> TikTok's practices have been criticized by a number of European thinktanks that have noted:
>
> - the temptation,in the last 2 years, of far-right parties to use TikTok as their main communication channel
>
> - lack of initiative from the platform to impose its own standards on political content, with more than 100 political advertisements identified in Germany in 2023;
>
> - the platform's lack of transparency in relation to researchers (access to data is through the **TikTok API, which provides less data than can be obtained through public** data scraping **methods**, and the **phenomenon of the source of funding for political content cannot be systematically analyzed** in the context **of the little information made available through the TikTok Ad Library).**
>
> *Thus EU Disinfolab considers **TtkTok's policy on the prevention of electoral misinformation as probably the least developed among social platforms**, and Democracy Reporting International has described **TtkTok's practices as a systemic risk to public discourse and electoral processes in the EU.***

- **without proper markings and outside the electoral campaign period**, aspects prohibited by Romanian electoral legislation.

As a consequence, **his visibility increased preferentially** in relation to the other candidates, whose posts were massively filtered, exponentially decreasing their online presence.

> According to the data obtained from TikTok, the application's algorithm **should normally have identified the coordinated nature** of Călin GEORGESCU's campaign and restricted the activity associated with the accounts, which did not materialize.

In the current electoral context, **an analysis of the online activities subsumed to Călin GEORGESCU's promotion campaign has been initiated at the TikTok level.**

> Tik TikTok's first report on the fact that a campaign to promote Călin GEORGESCU is being conducted in the TikTok environment was made in 2020. **In 2021 TikTok concluded that this was a suspicious activity.**

A massive promotional activity was also identified on the TikTok platform, **carried out in the last two weeks in support of POT (Young People's Party)**, a sovereignist party founded in 2023, which supports Călin GEORGESCU.

II. Financing mechanisms of Călin GEORGESCU's campaign

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Candidate Călin GEORGESCU declared to the Romanian electoral authority (AEP) 0 lei electoral campaign budget, i.e. that he did not spend any money during the campaign.

The held data revealed the involvement of the TikTok account Redacted used by **the Romanian citizen** **Redacted** in financing the promotion of Călin GEORGESCU on the socialization platform.

Using the Redacted account, he made donations on TikTok of over one million euros. After becoming the subject of public debate, the involvement of Redacted in financing the promotion of Călin GEORGESCU on the platform was confirmed by TikTok representatives in dialog with the Romanian authorities on 28.11.2024.

They confirmed the identity of the user Redacted and mentioned that he made payments in the amount of **381,000 dollars**, between 24.10-24.11.2024, **to users of TikTok accounts involved in the promotion of candidate Călin GEORGESCU, including after the date of the end of the electoral campaign. These represent violations of TikTok policies and Romanian electoral legislation.**

> **Redacted** displays **a standard of living that does not correspond** to the activities carried out through the company he owns, namely DIGITAL ON-RAMP SOLUTIONS SRL Brașov, established in 2021. Prior to the establishment of the company, he worked as a programmer in entities (**GLOBAYA FINTECH SRL and BITXATM TECHNOLOGY SRL**, both in Brașov) which are associated with the field of **cryptocurrencies** and are owned, among others, by **Redacted** (who is also a citizen of South Africa, where he emigrated in 1995).

Also, the financing of the TikTok influencers was ensured through the FameUp platform (dedicated to monetizing online promotional activities), where the advertising opportunity was published, along with well-defined descriptions. One of the methods of attracting Romanian influencers to collaborate in promoting **Călin GEORGESCU**'s candidacy was contacting them by e-mail by the South African company **FA Agency**, which offered them €1,000 for the distribution of a video clip made by them.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 238

ROMANIAN MINISTRY OF INTERNAL AFFAIRS

INFORMATION NOTE

i. After undergoing technical verifications on the internet and after analyzing metrical values identified on social media platforms, especially on Tik-Tok, in the context of elections and the electoral process, the existence of a electoral campaign that was not unclaimed by its actual authors, in which, beginning with November 2024, over 100 influencers (that could engage with roughly 8 million active followers) were subjected to manipulation activities targeted to the promoted candidate.

Tik-Tok did not implement the instructions of the Romanian Central Electoral Bureau regarding the marking as an electoral candidate of candidates that promote themselves on social-media platforms as well as the marking of electoral materials in general with the unique code that is provided by the Permanent Electoral Authority.

The metrics analysis showed a very sudden increase in the analyzed candidate (CĂLIN GORGESCU) between 13 and 26.11.2024, as he reached no. 9 in the global trend top according to the hashtags used by his campaign, however an actual artificial amplification campaign has not been identified on the Tik-Tok platform up until 24.11.2024, as the sudden surge of the views, which reached hundreds of millions was identified after 25.11.2024.

The analyzed data revealed approximately 130 Tik-Tok accounts which were used for disseminating video content with said content, using the hashtags *#echilibrusiverticalitate, #prezidentiale2024, #unliderpotrivitpentrumine (#balance and verticality; #presidential elections 2024; # the right leader for me),* the majority of these not being marked as paid publicity.

> *In this regard, videos were identified in which national influencers promoted a predefined scenario regarding elements that must be fulfilled by each candidate. The videos in question were posted on multiple platforms (TikTok, Instagram and Facebook), generating between 1.000 and 500.000 views per video, all of which had the hashtag #echilibrușiverticalitate*

The evaluation of the scenario in which the content elements were created, indicated similar situations with the ones carried out to influence the intention of Romanian influencers to promote the pro-Russian candidate from the Republic of Moldova in the analyzed posts.

Within the comments section related to each content element, a series of messages promoting a presidential candidate were identified. The analysis of these accounts indicates anomalies in their formulation, with indications regarding the use of fake accounts created only for the distribution of such comments.

Most influencers were not aware that they were promoting a specific candidate using *#echilibrusiverticalitate*. As of the date of this report only 51 of them still keeping their content elements associated with the three hashtags, the rest denouncing the messages and content as soon as they learned the identity of the candidate targeted by the respective electoral campaign.

An essential stage of the campaign was aimed at associating the mentioned ideal presidential profile with the candidate Călin GEORGESCU, through a high number of comments specifying his name, made for each promotional video post.

The influencers were co-opted and paid via a platform dedicated to micro-influencers.

At the level of the online marketplace through which the payment was processed, the request for the creation of publicity materials and services was submitted and published, along with clear instructions

regarding de promotion campaign, which included certain characteristics and hashtags corelated with the ideal type of presidents.

After the analysis of the content on the platform, it was revealed that the request for publicity services was posted by a certain company (possibly fake). More than that, influencers are paid for such campaigns depending on the number of followers, which in this case had as a calculation formula - 400 RON per 20.000 followers- the offer being accessible once every three days. Afterwards, the company could not be found anymore.

ii. Some of the campaign supporters involved in promoting and buying votes are exponents of right-wing extremist, criminal environments and religious cults, previously involved in promoting pro-Russian, anti-Semitic, anti-NATO or anti-Ukraine narratives, among which:

- One of them delt with the centralization of signatures (for Călin Georgescu) for running for president.

  He is a supporter of far-right ideologies, has been a fighter in the Foreign Legion. He openly promotes far-right ideologies by organizing para-military camps. Some of his objectives include the radicalization of youth, including minors, with the general purpose of the initialization of actions contrary to the rule of law.
- One of them is a known clan member and participated in a promotional campaign on Tik-Tok for Călin Georgescu.

  He participated in the said campaign by individual posts and live stories, where he openly associated with members of the organized crime world.
- Another two known members of the organized crime world posted content and participated financially in the campaign.

  On 23.11.2024 (one day before the election), he posted on his Facebook page a message in which he implied that the elections are being rigged. Currently, the post has a limited public exposure, however in one of the comments there was clear hate speech and instigation to public manifestations elements.

> *It must be mentioned that the account has several photos posted one day prior to the end of the campaign in which the independent candidate is promoted, while older posts show obvious promotion of the Russian president, Vladimir Putin. According to the Facebook profile, between 2018-2019, he mentioned he was in Moscow.*

Counterespionage activities revealed clear information that indicate the intentions of certain persons of undergoing actions that would affect the Romanian national sovereignty.

Following the analysis of the specific information space, in relation to the lessons learned and presented by other states, a similar action pattern was identified at the level of Ukraine, active in the period preceding the initiation of the aggression by the Russian Federation. Specifically, the information campaign "Balance and Quality" is identical to the "Brother next to Brother" campaign carried out by the Russian Federation in Ukraine, both being carried out through the manipulation of micro-influencers and algorithms.

As such, the campaign of influence of the Russian Federation in Ukraine was initiated via Telegram channels, organized by an anonymous Admin, who requested the creation of certain content elements on well-defined coordinates destined to promote the narrative of the cohesion of ethnical groups.

At the same time, the coordinators sent a set of minimum conditions that content creators had to respect (number of views, time coordinates at which the content elements had to be posted, the sound to be used, the emoticons to be used, the text-narrative). Also, within both campaigns, the initiators presented to the people involved, specific methods of avoiding the detection of content elements that do not comply with the policies of the TikTok platform, respectively ways to make the videos go viral.

The stated goal of the campaign was to raise awareness about the importance of participating in the voting process. Another similarity is the deletion of all data available online regarding the promotion campaigns to limit the possibility of attributing them to hostile actors. After contracting the content creation services, the influencers created and disseminated content elements according to the established coordinates, without knowing the real beneficiary of the promotion campaign. After the exposure of the influence campaign, several opinion vectors on social media criticized the actions of those involved, accusing them of promoting those hashtags only in exchange for financial benefits. In the Ukrainian campaign, each content element was paid 17USD/per content, while in Romania each influencer received 390 lei/ 20.000 followers. Also, the content creators involved in the contracted campaign in Ukraine were encouraged to distribute the content elements through other people, similar to the promotion campaign in Romania.

Exhibit 239

From: [redacted]@google.com]
Sent: 1/31/2024 10:54:03 PM
To: SMOLINSKA Joanna (EEAS-SAN FRANCISCO) [redacted]@eeas.europa.eu]
CC: Stephanie Seale (OHCHR Consultant) [redacted]@un.org]; SCHREURS Philip (EEAS-SAN FRANCISCO) [redacted]@eeas.europa.eu]
Subject: Re: EU/Norway panel on February 15
Attachments: image003.png; image003.png

Thanks for connecting us, Stephanie.

Hi Joanna! I hope all is well. Thanks for sharing the info below. [redacted] and I would recommend including [redacted] on the panel. She leads our global elections work for the YouTube public policy team. I've just confirmed with her that she can make the timing work.

Happy to connect on additional details as they come through.

Best,
[redacted]

On Wed, Jan 31, 2024, 4:14 PM SMOLINSKA Joanna (EEAS-SAN FRANCISCO) [redacted]@eeas.europa.eu> wrote:

> Thanks Stephanie for putting us in touch! I am putting Philip from our office in the loop as well as I will be off for meetings soon, so that the information flows.
>
> We are reaching out to Meta on the industry side as well [redacted] for that event. Please see below a fuller description. We would be thrilled to have [redacted] from YT but we understand that might not be possible. Could you let us know who else from YT (or yourself?) could step in? We would then check it back internally and revert.
>
> ***Brussels to the Bay event on Feb 15th 2024 @ 9 AM, 1 Post Street, SF:***
>
> ***"Strengthening democracy in a digital world: EU and UN policy tools to promote public discourse online"***
>
> 2024 is the biggest election year in world history. Countries making up more than half of the world population organize elections, including the European Union, the United States, India, and the United Kingdom. The internet plays an indispensable role in the democratic process, promoting the free flow of information and allowing voters to take part in the public debate. To strengthen our democracies, it is critical to safeguard people's fundamental rights, promote civic discourse and plurality on the internet, and boost trust in political processes. This Brussels to the Bay event, featuring special guest **Volker Türk (UN High Commissioner for Human Rights)**, will focus on the EU and the UN's efforts to reinforce democracy online.
>
> Both the EU and the UN have stepped up their efforts to strengthen democracy in the digital age. The EU, for example, adopted the Digital Services Act (DSA), a landmark regulation that aims to protect the rights of users

Confidential - Not For Public Release

by setting clear rules for very large online platforms to mitigate any risks to fundamental rights stemming from their services, coming into full effect on February 17. At the same time, efforts to address alleged impact of information in the digital space cannot be dissociated from the broader political and social context. Trust in political processes and democratic institutions also rests on how a State effectively recognizes and implements human rights offline and online and invests in the institutions needed for their protection.

What are the responsibilities of platforms in promoting public discourse? How can we leverage the DSA and other policy instruments to protect internet users' fundamental rights? What challenges does Generative AI present for the many elections ahead of us? Join us for this special Brussels to the Bay event, co-hosted by the EU Office in San Francisco and Norway, to find out.

Confirmed speakers

- Volker Türk (United Nations High Commissioner for Human Rights)
- Gerard de Graaf (Senior EU Envoy for Digital to the U.S. and Head of the EU Office in San Francisco)
- David Harris (Chancellor's Public Scholar at UC Berkeley, Senior Fellow at the Centre for International Governance Innovation, Senior Research Fellow at the International Computer Science Institute, and Visting Fellow at the Integrity Institute)
- Industry representative (Meta/ You Tube/ Microsoft tbc)
- Moderator: Alexis Wichowski (Professor of Professional Practice at Columbia University's School of Professional Studies, Program Director for the MS and Executive MS programs in applied technologies / technology management, and affiliate faculty with Columbia's School of International & Public Affairs)

Agenda (to be confirmed and subject to changes)

- 09:00 – 09:15 | Walk-in
- 09:15 – 09:20 | Introductory remarks by Gerard de Graaf
- 09:20 – 10:45 | Panel discussion with Volker Türk, Gerard de Graaf, David Harris, Meta/ YT/ MSFT (tbc) moderated by Alexis Wichowski

Registration

In-person + on Zoom

Confidential - Not For Public Release

From: Stephanie Seale (OHCHR Consultant) <[redacted]@un.org>
Sent: 31 January 2024 13:09
To: [redacted] [redacted]@google.com>; SMOLINSKA Joanna (EEAS-SAN FRANCISCO) [redacted]@eeas.europa.eu>
Subject: EU/Norway panel on February 15

Dear [redacted]

With this email I wanted to put you in touch with Joanna Smolinska, who is the fabulous deputy head of office for the EU here in San Francisco. She is organizing the panel I mentioned earlier with Norway on "Strengthening democracy in a world of social media and AI." I've told her that [redacted] is unavailable that week, so thought we could triangulate the conversation here and see whether there might be a suitable replacement from YouTube or Google who could speak in [redacted] stead. I don't know who else from industry has been invited but Joanna is running the show so she is the best resource here.

Thanks!

Stephanie.

Stephanie Linden Seale
Outreach and Partnerships, Tech & Human Rights
Office of the United Nations High Commissioner for Human Rights
E-mail: [redacted]@un.org *and* [redacted]@gmail.com
Mobile: +1 6[redacted]
Web: https://www.ohchr.org/en/topic/digital-space-and-human-rights

Confidential - Not For Public Release

Exhibit 240

From: [redacted]@google.com]
Sent: 2/1/2024 2:04:38 PM
To: [redacted]@google.com]
Subject: Fwd: EU/UN panel on February 15
Attachments: image002.png

who do you think could do this?
if we get [redacted] that would be more human rights than elections

---------- Forwarded message ---------
From: [redacted] <[redacted]@google.com>
Date: Wed, Jan 31, 2024 at 11:17 PM
Subject: Fwd: EU/UN panel on February 15
To: [redacted] <[redacted]@google.com>, [redacted] <[redacted]@google.com>, [redacted] <[redacted]@google.com>

Hi team,
Another pitch by the EU in SF team to have [redacted] at an upcoming conference. I feel like we can't decline their invites forever ;-) If [redacted] isn't a good candidate for this once, would there be anyone else interested to speak obo YT/ Google?

Thanks,
[redacted]

---------- Forwarded message ---------
From: **SMOLINSKA Joanna (EEAS-SAN FRANCISCO)** <[redacted]@eeas.europa.eu>
Date: Wed, Jan 31, 2024 at 11:36 PM
Subject: EU/UN panel on February 15
To: [redacted] <[redacted]@google.com>
Cc: SCHREURS Philip (EEAS-SAN FRANCISCO) <[redacted]@eeas.europa.eu>

Hi [redacted]

I hope you are well!

Looping you in – we are organising an EU-UN event on Feb 15th, see details and an email to Alex below.

Can you let me know if [redacted] is indeed unavailable (UN people have be in touch with her team)? I can shoot an official email if that helps.

Joanna

Confidential - Not For Public Release

From: SMOLINSKA Joanna (EEAS-SAN FRANCISCO)
Sent: 31 January 2024 13:14
To: 'Stephanie Seale (OHCHR Consultant)' <[redacted]@un.org>; [redacted] <[redacted]@google.com>
Cc: SCHREURS Philip (EEAS-SAN FRANCISCO) [redacted]@eeas.europa.eu>
Subject: RE: EU/Norway panel on February 15

Thanks Stephanie for putting us in touch! I am putting Philip from our office in the loop as well as I will be off for meetings soon, so that the information flows.

We are reaching out to Meta on the industry side as well ([redacted]) for that event. Please see below a fuller description. We would be thrilled to have Leslie from YT but we understand that might not be possible. Could you let us know who else from YT (or yourself?) could step in? We would then check it back internally and revert.

Brussels to the Bay event on Feb 15th 2024 @ 9 AM, 1 Post Street, SF:

"Strengthening democracy in a digital world: EU and UN policy tools to promote public discourse online"

2024 is the biggest election year in world history. Countries making up more than half of the world population organize elections, including the European Union, the United States, India, and the United Kingdom. The internet plays an indispensable role in the democratic process, promoting the free flow of information and allowing voters to take part in the public debate. To strengthen our democracies, it is critical to safeguard people's fundamental rights, promote civic discourse and plurality on the internet, and boost trust in political processes. This Brussels to the Bay event, featuring special guest **Volker Türk (UN High Commissioner for Human Rights)**, will focus on the EU and the UN's efforts to reinforce democracy online.

Both the EU and the UN have stepped up their efforts to strengthen democracy in the digital age. The EU, for example, adopted the Digital Services Act (DSA), a landmark regulation that aims to protect the rights of users by setting clear rules for very large online platforms to mitigate any risks to fundamental rights stemming from their services, coming into full effect on February 17. At the same time, efforts to address alleged impact of information in the digital space cannot be dissociated from the broader political and social context. Trust in political processes and democratic institutions also rests on how a State effectively recognizes and implements human rights offline and online and invests in the institutions needed for their protection.

What are the responsibilities of platforms in promoting public discourse? How can we leverage the DSA and other policy instruments to protect internet users' fundamental rights? What challenges does Generative AI

Confidential - Not For Public Release

present for the many elections ahead of us? Join us for this special Brussels to the Bay event, co-hosted by the EU Office in San Francisco and Norway, to find out.

Confirmed speakers

- Volker Türk (United Nations High Commissioner for Human Rights)
- Gerard de Graaf (Senior EU Envoy for Digital to the U.S. and Head of the EU Office in San Francisco)
- David Harris (Chancellor's Public Scholar at UC Berkeley, Senior Fellow at the Centre for International Governance Innovation, Senior Research Fellow at the International Computer Science Institute, and Visting Fellow at the Integrity Institute)
- Industry representative (Meta/ You Tube/ Microsoft tbc)
- Moderator: Alexis Wichowski (Professor of Professional Practice at Columbia University's School of Professional Studies, Program Director for the MS and Executive MS programs in applied technologies / technology management, and affiliate faculty with Columbia's School of International & Public Affairs)

Agenda (to be confirmed and subject to changes)

- 09:00 – 09:15 | Walk-in
- 09:15 – 09:20 | Introductory remarks by Gerard de Graaf
- 09:20 – 10:45 | Panel discussion with Volker Türk, Gerard de Graaf, David Harris, Meta/ YT/ MSFT (tbc) moderated by Alexis Wichowski

Registration

In-person + on Zoom

From: Stephanie Seale (OHCHR Consultant) [redacted]@un.org>
Sent: 31 January 2024 13:09
To: [redacted] [redacted]@google.com>; SMOLINSKA Joanna (EEAS-SAN FRANCISCO) <[redacted]@eeas.europa.eu>
Subject: EU/Norway panel on February 15

Dear Alex,

Confidential - Not For Public Release

With this email I wanted to put you in touch with Joanna Smolinska, who is the fabulous deputy head of office for the EU here in San Francisco. She is organizing the panel I mentioned earlier with Norway on "Strengthening democracy in a world of social media and AI." I've told her that [redacted] is unavailable that week, so thought we could triangulate the conversation here and see whether there might be a suitable replacement from YouTube or Google who could speak in [redacted] stead. I don't know who else from industry has been invited but Joanna is running the show so she is the best resource here.

Thanks!

Stephanie.

Stephanie Linden Seale
Outreach and Partnerships, Tech & Human Rights
Office of the United Nations High Commissioner for Human Rights
E-mail: [redacted]@un.org *and* [redacted]@gmail.com
Mobile: +[redacted]
Web: https://www.ohchr.org/en/topic/digital-space-and-human-rights

--

[redacted] | Governmental Affairs & Public Policy | [redacted]@google.com | +[redacted]

This email may be confidential or privileged. If you received this communication by mistake, please don't forward it to anyone else, please erase all copies and attachments, and please let me know that it went to the wrong person. Thanks.

Confidential - Not For Public Release

Exhibit 241

Message

From: [redacted] [redacted]@meta.com]
Sent: 2/26/2024 6:19:40 AM
To: Meta DSA Compliance Office [redacted]@meta.com]; DSA Inbox Group [redacted]@meta.com]
CC: [redacted] [redacted]@meta.com]; [redacted] [redacted]@meta.com]
Subject: Re: Roundtable Invite DSA Election Guidelines

Thank you
I got this invitation as well.
Best regards

Meta
[redacted]
Director and Deputy Head of EU Affairs

35 Boulevard du Régent
B-1000 Brussels

M: +[redacted]
E: [redacted]@meta.com

Assistant: [redacted]
E: [redacted]@meta.com

We are contacting you in connection with policy communications and related engagements at this email address because you previously provided your contact information to the Meta team or because it was publicly available in your official capacity. For further detail on how we process and store your data (including how we may record our engagements with you) and to learn more about how we safeguard data, please see our privacy notice: https://fb.me/PolicyCommunicationsPrivacyStatement. Meta takes the protection of your data seriously. If you do not wish to receive our policy communications, please let us know and we can unsubscribe you.

Produced to HJC

From: Meta DSA Compliance Office <[redacted]@meta.com>
Sent: Monday, February 26, 2024 2:52 PM
To: DSA Inbox Group <[redacted]@meta.com>
Cc: [redacted] <[redacted]@meta.com>; [redacted] <[redacted]@meta.com>
Subject: Fw: Roundtable Invite DSA Election Guidelines

Invitation from the EC for a RT on the draft guidelines for election integrity.

The XFN on elections meets this afternoon and we will get this on the agenda to discuss.

[redacted]
Senior Compliance Manager
10 Brock Street, London NW1 3FG
Meta | Mobile +[redacted]

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

From: CNECT-DIGITAL-SERVICES@ec.europa.eu <CNECT-DIGITAL-SERVICES@ec.europa.eu>
Sent: Monday, February 26, 2024 1:33 PM
Subject: Roundtable Invite DSA Election Guidelines

Dear all,

As you may be aware, under the Digital Services Act the Commission has been actively working on the topic of election integrity. The Commission has engaged with several providers of VLOPs and VLOSEs in readiness dialogues on the topic in the run up to previous Member State elections and is planning to have a continuous structured election dialogue with relevant designated VLOPs and VLOSEs on preparedness for the upcoming Member State and European Parliament Elections.

In this context, we are pleased to invite you to a roundtable on the topic of election integrity under the DSA. The roundtable will be held **online** and take place on **Friday 1st March from 14:30-16:30.**

The aim of the roundtable is to present and answer any questions you may have on the **draft DSA guidelines for election integrity** published by the Commission in line with Article 35(3), to present best practices and recommend possible mitigation measures. We also would like to offer you the opportunity to present evidence based best practices of mitigation measures you employ in the area of elections.

In addition, we welcome your feedback on the draft guidelines directly through the public consultation available here.

We would appreciate you letting us know who will join from your organisation, and if you want to use the opportunity to present any evidence based best practices by emailing our secretariat at CNECT-DIGITAL-SERVICES@ec.europa.eu

We will send a link to join the roundtable closer to the date of the event.

Best regards,

The DSA Team of the Commission

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 242

From: GEESE Alexandra [redacted]@europarl.europa.eu]
Sent: 2/29/2024 9:59:58 AM
To: undisclosed-recipients:
Subject: Hearing with Frances Haugen & Maria Ressa 6/3, 15-18:00, EU Parliament, JAN 6Q2: PROTECTING THE 2024 ELECTIONS: FROM ALARM TO ACTION

PROTECTING THE 2024 ELECTIONS: FROM ALARM TO ACTION

HOW THE DSA CAN PROTECT DEMOCRACY BY MEANINGFULLY ADDRESSING THE VIRALITY OF DISINFORMATION AND HATE SPEECH

Societies across Europe are becoming more polarised, and widespread disinformation undermines trust in democratic institutions and electoral processes. Online platforms have spurred the proliferation of conspiracy theories, and disinformation, and actively encouraged toxic polarization as an anti-democratic instrument. With the widespread accessibility of large language models, this phenomenon is set to be supercharged. In light of the looming risks facing the 2024 European elections, rigorous enforcement of the Digital Services Act (DSA) is crucial – and with the integrity of European democracy at stake, the responsibility is monumental.

This event will explore electoral integrity in the online information space with a particular focus on engagement-based ranking and incentivising the production of divisive content. How can DSA implementation support the integrity of elections? What can platforms do to stop harmful amplification? How do less toxic recommender systems look like and what are the alternatives for content ranking on social media platforms? What can the European Commission and enforcement agencies do to make sure that elections take place in a safe online environment?

This Hearing will be co-chaired by MEPs Katalin Cseh / Renew Europe, Alexandra Geese / Greens/EFA and Paul Tang / Progressive Alliance of Socialists and Democrats. It will be livestreamed on YouTube and followed by some drinks and snacks.

Room JAN 6Q2
European Parliament, Brussels

Confidential - Not For Public Release

CLICK HERE TO REGISTER

WITH A KEYNOTE BY:

- Frances Haugen, Former Facebook Employee & Whistleblower; Advocate for Public Oversight of Social Media (remote)

…AND CONCLUDING REMARKS BY:

- Maria Ressa, CEO of Rappler, Journalist, Author, Freedom Advocate and 2021 Nobel Peace Prize Winner (in person)

1. Panel: recommender systems, disinformation, elections and the DSA

Speakers:

- Renate Nikolay, Directorate-General for Communications Networks, Content and Technology, EU Commission
- Marc Faddoul, transdisciplinary technologist, expert on recommender systems and algorithmic audit, Director and Co-Founder of AI-Forensics
- Tanya O'Carroll, independent advisor, strategist and leader focused on tech accountability, human rights and social justice

2. Panel: CSO meet enforcement authorities roundtable - the future of recommender systems

Speakers:

- Katarzyna Szymielewicz, Co-Founder and President of the Panoptykon Foundation
- Arturo Bejar, Whistleblower, former Facebook employee and consultant for Instagram
- Dr. Martin Degeling, Technical Expert "Auditing of Recommendation Systems" at the Stiftung Neue Verantwortung
- Felix Kröner, Policy Advisor at RESET.tech
- Emily Gorcenski, Data Scientist, Thoughtworks
- Anne Marie Pollock, Director of Policy, “Democracy and Fundamental Rights”, Coimisiún na Meán
- Dr. Julia Marquier, Head of Unit “Net neutrality, Platforms and Artificial Intelligence” at the German enforcement authority, Bundesnetzagentur
- Prabhat Argwal, Head of Unit of the 'Digital Services and Platforms' Unit at the European Commission's DG CONNECT

Confidential - Not For Public Release

Exhibit 243

EC ROUNDTABLE FRIDAY 1 MARCH RE: DRAFT DSA ELECTIONS GUIDELINES

VIRTUALLY, 14:30-16:30

Invitation: *As you may be aware, under the Digital Services Act the Commission has been actively working on the topic of election integrity. The Commission has engaged with several providers of VLOPs and VLOSEs in readiness dialogues on the topic in the run up to previous Member State elections and is planning to have a continuous structured election dialogue with relevant designated VLOPs and VLOSEs on preparedness for the upcoming Member State and European Parliament Elections.*

In this context, we are pleased to invite you to a roundtable on the topic of election integrity under the DSA. The roundtable will be held ***online*** *and take place on* ***Friday 1st March from 14:30-16:30.***

The aim of the roundtable is to present and answer any questions you may have on the ***draft DSA guidelines for election integrity*** *published by the Commission in line with Article 35(3), to present best practices and recommend possible mitigation measures. We also would like to offer you the opportunity to present evidence based best practices of mitigation measures you employ in the area of elections.*

HIGH LEVEL REACTIONS IF NEEDED

- Welcome the EC opening a consultation on these guidelines and opportunity (for industry and all other stakeholders) to provide reactions.
- Encouraged to see the draft draws from dialogue with industry and existing multi-stakeholder frameworks like the EU Disinfo Code.
- Still studying the guidelines, planning on submitting a response to the consultation, and would welcome more dialogue and exchange.

QUESTIONS IF OPPORTUNITY

1. **Risk Profile:** we assume that all of the suggested measures are not intended to apply to each election in region, but rather should map to relevant risk profile. Can you tell us more about the reference to the risk profile in Section 3 of the guidance and how this is intended to work?

Prabhat's answer:

- *Guidelines are the EC's opinion on what are good mitigation measures to comply with DSA. Not an Implementing Act, it is a guidance document: do not have to follow all of them. However, obligation is to assess and mitigate systemic risks. Those are best practices to do that.*
- *Onus if deviate from this to have alternative measures that are equal or better than those in here. Can be adapted, and recognise that not every regulated entity has the same risk profile. Not every platform in the list of VLOPs has an election risk - some do not.*
    - *Gave Snap as an example of a platform that has a different type of service and features which means they will need to adapt the Guidelines somewhat. Referred to the prominence of the map feature on Snap as an (illustrative only) example of something that differentiates them from others and therefore means Guidelines implementation may be different for them.*

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED - NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

- *Not all comprehensively binding, but catalogue of best practices that need to adapt and where deviate, important to document deviation and able to answer questions on that to the EC. They've been designed based on what understood/learned from working with platforms and civil society. Flip side: If some of these don't work, also want to be able to demonstrate that.*
- *Careful what elements pick/choose and what elements are disregarded.*

2. **On the rapid alert system:** could you provide us with additional context on the concern that you are looking to solve for, and how you see this interacting with DSA obligations to have a trusted flagger system?
3. **Country input:** To what extent do these guidelines reflect input from Member States? The section of the guidelines focused on cooperation with national authorities presupposes commitment and resources, it would be interesting to understand their view on what is feasible and most impactful without asking too much of them.

Prabhat's answer: Dedicated meeting of the Board next week on the guidelines. Many Member States insisted wanting a rapid alert system. Seeking active feedback, and can provide this input (including concerns over duplication) in answer to consultation. Objective of EC is to ensure clear escalation paths.

4. **On the Generative AI recommendations:** could you provide clarity on the expectation on providers with adopting these measures given the challenging landscape of detecting and identifying AI content, and cross-industry efforts to address deceptive AI risks (e.g., Munich Security Accord)?

Microsoft and Snap mentioned some of the changes on the Gen-AI recommendations, including detection. EC took note and encouraged this to be captured in consultation response.

Prabhat: What happens if a non-watermarked ai-generated image of a president or political leader circulates during an election. We want to make sure you are prepared. And we think that there will be residual risk there that an escalation/rapid alert system with the EC and DSCs allows us to address. Noted considering table top exercises on this with industry. [redacted] flagged our interest in understanding more about such an exercise. Prabhat welcomed that reaction and said this was a very early idea and he may follow up.

5. **Code/Guidelines:** How are you thinking about the interaction between the DSA, the Codes (Disinfo and Hate Speech) and the Guidelines? We see quite a lot of overlap, which as a signatory we find encouraging to the extent that the measures are consistent - but we wonder how these various instruments will interact in a practical sense, including how the EC will maintain proportionality of its expectations towards mitigation measures
6. **On the performance metrics recommendations:** we assume that these recommendations are meant to be flexible so that each platform can identify the metrics that are most pertinent to them?
7. **Next steps:** How do you expect the process to go after the closing of the consultation?
8. Are there other areas of DSA where you are expecting/planning on publishing this type of consultation?

Pinterest: No comments on reactions to platforms, because election/political content not a significant use case on Pinterest. Ban political ads. And simply take down concerning content rather than using things like labels/media literacy.

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Light touch summary of election approach from Twitter.

Prabhat also flagged that they are thinking further about how to provide input on the recommender system aspects of the products.

Overall: Prabhat encouraged platforms to include our best practices in our consultation response.

REMARKS ON BEST PRACTICES

- We learned from working on more than 200 elections around the world since 2016.
- For every election, we evaluate whether our standard mechanisms (which include the policies, tools and processes that we document in our EU Code of Practice on Disinformation reports) address the election-specific threats for that election. Should there be any outlying risks, we work across several teams and sometimes with external partners to put appropriate measures in place to mitigate those risks.

5 pillars:

1. **We have advanced security operations** to take down manipulation campaigns and identify emerging threats. We also continue to carry out proactive sweeps on the platform to catch any banned organizations attempting to violate our policies or cause offline harm.

2. **We remove harmful content: We remove content likely to directly contribute to a risk of interference with people's ability to participate in elections** (i.e. content with false information on how or where to vote).

3. **We connect people with reliable information about voting while combating misinformation across languages.**
    - *Access to official information*: We connect people with details about voter registration and the election through in-app notifications.
    - *Cooperation with fact-checking organisations and contextual information*: We have the largest fact-checking network of any provider, including 26 independent fact-checking organisations across the EU covering 22 languages. When content is debunked by these fact checkers, we attach warning labels to the content and reduce its distribution in-feed so people are less likely to see it.

4. **We provide industry-leading transparency for ads about social issues, elections or politics,** and have built the largest global fact-checking network of any platform. We also reject ads encouraging people not to vote or calling into question the legitimacy of the upcoming election.
    - Advertisers who run ads about social issues, elections, or politics are required to complete an authorization process and include a "paid for by" disclaimer on these ads.
    - We provide information about advertiser targeting choices, and ads delivery, in the publicly available Ad Library.
    - [We are working on implementing the recent EU Regulation on transparency in political ads in this space as well.]

Provided to HJC

CONFIDENTIAL TREATMENT REQUESTED - NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

5. **Civic engagement:**
    - *Media literacy*: We run targeted media literacy campaigns as part of our work to combat misinformation. For the EU elections, are working with the European Fact-Checkers Standards Network and the European Disability Forum on 2 campaigns.
    - For national elections in the EU, we typically launch feed notifications to connect people with election information.
    - We also conduct training sessions to ensure that **candidates know how to use the various tools and features available on our platforms** to protect their accounts from issues ranging from hacking attempts to harmful content.

READOUT MARCH 1 VIRTUAL MEETING WITH DSA UNIT

Nature of the meeting: the EC called a number of VLOPs to discuss the draft DSA guidelines on elections, gather questions and exchange best practices.

On March 1, the EC made a strong case for companies to meet their expectations with respect to effective mitigation measures in the area of election integrity. The EC believes in flexibility, but they expect VLOPs to assess all elections. Overall, the EC wants to explore the topic of Gen AI and elections and how our recommender systems work with respect to political content. Some additional, detailed observations follow.

Main observations

- No level playing field across industry. The EC Meeting today gathered both VLOPs that are part of the Code of Conduct on Disinformation and those who are not; there is no level playing field currently around measures foreseen to address the expectations of the regulator on DSA elections integrity. Examples are Pinterest or Snap versus Google, LinkedIn and Meta. This made this meeting less effective.
- The guidelines are **not an implementing act (**where we have to comply with every measure**)** but if we are not taking on board a measure they recommend, we are expected to document the reasons why and how we mitigate the said risks. The Guidelines are not a Communication, they are regulatory guidance.
- EC has strong views about the DSA legal requirement to involve **external stakeholders** in the evaluation/development of our measures. This was sparked by MSFT questions around external red teaming requirements. The EC relied on Recital 90 of the DSA which they stressed as the legal basis for our obligation to put mitigation measures in place together with external partners.
- EC is working with the Digital Service Coordinators (DSC) Board on a **Rapid Alert System (RAS).** Several Member States are pushing for such a system. This confirms the information we have about the DSC focus on election integrity. The DSCs have created a WG on Democracy including electoral authorities as well. The EC is holding a meeting with them on the Guidelines specifically to include their feedback. They also explained that their priority is an escalation path, which in their view is necessary to address "residual risk" during elections. The focus is on authorities rather than civil society, but we cannot discard the involvement of the latter. In a preliminary analysis, a RAS with authorities is preferable to an ad hoc one with civil society outside the DSA trusted flagger program.
- The EC is thinking of organizing a **tabletop exercise** on incident response with the companies, to focus on the practicalities and avoid dictating new rules that might not be

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

operational or effective. The example they raised was platforms' readiness to deal with an AI-generated image or video of a politician doing or saying something that they did not in fact do/say that could impact elections.

- DSA guidance on elections integrity does not end with these guidelines. The EC intends to publish the final guidance by the end of this month or beginning of April, which suggests *they are looking for validation, not revision of what they already have* and or complying with their obligation to consult externally. The EC indicated they will review the guidelines next year.

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 244

From: Guénolé Carré [redacted]e@political-intelligence.com]
Sent: 3/8/2024 9:03:22 AM
To: [redacted] [redacted]@google.com]
CC: Google [redacted]@political-intelligence.com]
Subject: RE: Hearing with Frances Haugen & Maria Ressa 6/3, 15-18:00, EU Parliament, JAN 6Q2: PROTECTING THE 2024 ELECTIONS: FROM ALARM TO ACTION

Dear [redacted]

Please find below a summary and a readout of the event organized by the Greens titled "Protecting The 2024 Elections: From Alarm To Action", which took place earlier this week.

Summary:

- A keynote speech by whistleblower Frances Haugen highlighted the need to scrutinize platforms intensely as they are being weaponized. She also noted the new paradigm of adversarial AI, which complicates once more the task of identifying illegal content or disinformation online.
- A first panel saw Renate Nikolay (DG CNECT), Marc Faddoul (AI-Forensics) and Tanya O'Carroll (former Amnesty International) discuss the impact of recommender systems, especially in the context of election and disinformation.
- Overall, the tone was fairly critical of the current state of play with recommender systems, but a strong emphasis was put on making sure to implement the DSA fully, possibly with additional guidelines for platforms, and not so much on additional legislation for the next mandate. All panellists praised the DSA as being a first of a kind legislation, but they all warned that it might not be sufficient in the long run or if poorly implemented.
- Asked separately about the impact of violence against women online, Renate Nikolay expressed bitterness as to how little advances had happened during this mandate on VAW, and she noted that it was one of the systemic risks included in the DSA for VLOPs. Answering a separate question, she noted that the Commission was very diligent regarding the independence of DSC, including in countries such as Hungary. She also emphasised on the role of the guidelines currently being finalised regarding the elections.
- A second panel saw civil society activists discuss with enforcement authorities on the future of recommender systems.
- Multiple civil society activists promoted third party recommender systems as a possible alternative, called for platforms to make flagging content more accessible and actionable, promote a different set of values within their systems (putting profit and ads lower in the promotional aspect). All called for detailed analysis of platforms and their different business models, and for stronger enforcement.
- Enforcement authorities, including Irish, German and European enforcers, said that they are making the most out of these new investigative powers, detailing ongoing procedures and investigations, but none make forward looking statements on where they see evolutions regarding recommender systems. They all called for more data from companies, and for civil society to interact with them with evidence.

Please let us know if you have any questions or comments.

Best regards,

Guénolé

Readout:

Panel 1 : Recommender systems, disinformation, elections and the DSA

Speakers:

- Renate Nikolay, Directorate-General for Communications Networks, Content and Technology, EU Commission
- Marc Faddoul, transdisciplinary technologist, expert on recommender systems and algorithmic audit, Director and Co-Founder of AI-Forensics
- Tanya O'Carroll, independent advisor, strategist and leader focused on tech accountability, human rights and social justice

Confidential - Not For Public Release

Alexandra Geese (Greens):

- The elections are threatened by Russians. The majority of people does not dare to express their political opinion on the internet and 25% withdraws from the public discussion. Hate pushes people to violence, also in the real world. There is a huge amount of disinformation and emotional content available online. I believe this is a threat to democracy and it is important we tackle it.
- We thought we could tackle this in the DSA by putting transparency into recommender systems. The DSA gives so many possibilities to the Commission, but also to us and CSOs. We should focus a bit more on recommender systems, and on how content is distributed.
- Let's listen to our keynote speaker Frances Haugen. She is known as the Facebook whistleblower. The Wall Street Journal published articles based on her information. Her testimonial in this EP in the past really made a difference, and helped to get comprehensive transparency in the DSA.

Keynote speech Frances Haugen:

- We are looking at an inflection point, because half of the people in the world will be voting soon. There are dangers posed by systems with very few checks and balances which are made for private gains.
- The products we use online today are social platforms. Children spend on an average 3,5 hours a day on them. It is very different from what we used in 2010. Back then we used it much more like our email inboxes. We chose our friends/pages we followed ourselves. Today the majority of content we see on Facebook and Instagram are not shown because we chose it, but because an algorithm chose it. Our relationships we trust, we are delegating that control to computers, and we have little insight on how they operate.
- Algorithms are optimized in showing to us "what looks like good". However, "good" looks like more clicks, more, more time spent, more ads clicked on. And what bleeds leads. It is shown that the shortest path to a click is anger.
- Facebook is keeping us engaged as long as possible, which means also misinformation was going up, and the amount of hate speech was going up as well. This is because of our vulnerability. Our engagement is based on extreme content, because we are drawn to extreme content.
- **The vulnerability of these platform to being weaponized.** This is why it is vitally important to use the DSA pathways for oversight. **We have to use the powers under DSA to look how vulnerable a platform is to being exploited by algorithms.** Because of things like targeted advertisement for example you can get engagement by true facts.
- **The strategy of content moderation as a way to create safety is over.** We have to look at what content is distributed on the platforms.
- Second thing is that there is a new paradigm: **adversarial AI.** This is a field of study that has been around for years. In a world where an adversarial AI can take something down, they can try to train AI to change something slightly to e.g. an image, so it can get through the AI security check anyways.
- Half the world is voting. The DSA and EU's actions for oversight are very important, as you are the most robust framework in the world. I hope you will strengthen the DSA and the actions the COM is taking. We need the EU to protect democracy.

Alexandra Geese (Greens):

- What can we actually do? Where should our priorities lay. You mentioned that democracies are getting fragile. It is true, we have seen how quickly political developments can go, also within the EU.
- Let's introduce the first panel now.

Katalin Cseh (Renew):

- I am an MEP from Hungary, where the government has become the biggest spreader of misinformation itself and we have seen the impact on democracy, but it is a global issue. Social Media platforms can give malign actors the power to undermine democracy.
- Online platforms, if they are not regulated properly, they can be used to spread disinformation.
- I am very proud to be an MEP in a Parliament that has voted for the DSA. This is such a good solid foundation.
- I find transparency requirements the most important: we have to remember that democracy also means a meaningful public debate and a well-informed electorate.

Renate Nikolay, Directorate-General for Communications Networks, Content and Technology, EU Commission

- It was already a theme during the elections of 2019. This year 2024 we are much better equipped to use our tools in place.
- We are also looking into on cybersecurity and its impact on elections. A lot of threats are cyberthreats and they include disinformation influenced by foreign powers.
- We talked about the DSA, but let's not forget the DMA. Tomorrow is the deadline for the compliance report of 6 gatekeepers.
- **The DSA is a holistic approach, it is not going after one incident, but it is systemic. How are platforms using the space and how we can push them to do better? We have already opened 2 investigations into X and**

Confidential – Not For Public Release

TikTok. We were alone first, because we had to deal with the VLOPs. We already see effects from all that work, we see more transparency, and we see that it is easier to flag illegal content, easier to appeal against content moderation decisions, and there is control of recommender systems, but it is not enough. We have to go further.

- **We also work on useful input such as from civil society experts.** We use legal transparency rules on recommender systems. We also do risk assessment that shows design of recommender systems.
- **Transparency is key**. We already have the transparency data base with billions of entries, which shows content moderation platforms have already done a lot. We also have the data access. X and TikTok investigations also target recommender systems, addictive design, etc. We are also looking at generative AI.
- Finally we are building on a tool in the DSA. **We are preparing guidelines ahead of the elections**, so platforms will have enough time for higher due diligence before elections. We want them to watermark deepfakes, etc. It is really on our radar and cannot be perfect after a few months already, but we will keep pushing them.

Marc Faddoul, transdisciplinary technologist, expert on recommender systems and algorithmic audit, Director and Co-Founder of AI-Forensics

- There is this intersection of AI and disinformation. **It is important to know what content gets distributed**. There is a lot of focus on the production side right now, but we should not shift our attention away too fast to the distribution side of it.
- Most of the content online is recommended by an algorithm, which is designed to keep you engaged. And this content tends to be the most sensational and does not provide a level playing field for political discourse.
    - For example, ahead of the elections in France Eric Zemmour was getting 30 % of engagement on platforms, while he eventually only got 7% of the votes. And this risk is still occurring right now.
    - There is also an increasing reliance on influencers regarding political discourse. This is on the distribution side.
- **There are also new risks on Gen AI side; the creation of forms of disinformation, such as deepfakes, fake news websites, … Gen AI can also enable troll armies that feed social media with comments and push content.**
- And then you also have this phenomenon of misinformation, which is when AI unintentionally produces disinformation.
    - For example, ahead of the German elections, when we asked chatbots about election information (who is running, etc.), in 30 % of the cases the bot was giving information with errors. There is a structural problem, and Microsoft would fix the problem, but today there are still the same tendencies. This is a reminder that self-regulation is something we cannot count on. Now we luckily have strong regulations (DSA and DMA), but now we have to focus on enforcement.

Tanya O'Carroll, independent advisor, strategist and leader focused on tech accountability, human rights and social justice

- The past 6 months we have seen fake political scandals, fake voice recordings, deepfakes and so on. These new challenges come on top of everything else we have not fixed yet.
- Many of the corporations have dismantled their content moderation teams, which is also a sign of their lack of interest in that issue.
- **It is only here in Europe that we have this powerful toolbox (DSA) to tackle this problem. The measures that are in the COM guidelines on elections are very strong messages. But there is currently nothing in the guidelines that really covers recommender systems. We know it is the worst content that is pushed forward, because it keeps people engaged.**
- The good news is that we still have 3 months to go before the elections, but there is much to be done. Platforms could make a switch and turn off the toxic version of the recommender system. This would automatically help the problem. This needs to be put in place.

Q&A:

Q by Global Disinformation Index: How do you feel about third-party quality assessments to break the link of the content being the most engaged with, also being the most pushed content?

Q: Would opening recommender systems to third parties lead to more choice? How will you respond to a crisis occurring during the elections?

- **Renate Nikolay**: There is a crisis mechanism in the DSA but this is only for severe crises. We created an incident response mechanism. An incident can be local, regional, etc., but we should still be able to react fast and this is the idea of the guidelines.
- **Marc Faddoul**: An algorithm will not be the solution for quality content. We need humans. To do third party assessment, use organizations, and give agency back to the users to choose what they want to look at.
- **Renate Nikolay**: the transparency power of the DSA will increase the momentum of third parties, who could deal with quality. It could also be a task for the research/ academic world where we have powerful information on

Confidential - Not For Public Release

addiction/mental health. It is not black and white. I do not believe the telecoms experience is the best example of how to go about this because this would create 27 markets. We want to shake up these business models.

Q by Tanya O'Carroll: Ms. Nikolay, I am happy to hear that recommender systems are on your radar and that you are asking the tough questions, but the 14 week clock before the elections is ticking right now. We know that platforms can substantially change their recommender systems (eg. US elections). They have done it before, and they can do it again. To what extent can we use the DSA to push these changes in these 14 weeks?

Q by Tanya O'Carroll: How do we move from algorithmic systems, what are you opting for instead? Who should be deciding those things? We don't want the platform to decide what is quality content, we need an independent body to decide this. We need to increase user empowerment and choice, and the EP did an excellent job this year. There are certainly some interesting things to build on in the next mandate.

Q: What is the difference between political and non-political information?:

- **Renate Nikolay**:
  - I think it goes wider in terms of scope. Not all platforms are the same. When you look at our 40 RFIs, they are very different. It is the first time we got a risk assessment of the VLOPS, and they don't have the same structure, so this is something to be worked on in the future.
  - The guidelines highlight the importance of the elections risk, and they have to take into account diversity. I think we have the recommender system very much on our radar in the guidelines. The idea is to make sure that there is no automatic distribution that goes on forever.
  - In each of the national elections we did talk to the platforms, NGO's,… and we talked to them afterwards as well to draw lessons.

Q by Alexandra Geese (Greens): How can the guidelines help for the elections you think?

- **Renate Nikolay**: We have the DSA, but will that prevent everything? We are not sure, but it will be a push (e.g. watermark) and maybe it will make them act faster than in the Slovakian case. There is an awareness we want to create as well.

Q: Ms. Nikolay, mindful of the 14 weeks before the elections, have you had answers to the RFIs that were sent?

Q by Katalin Cseh (Renew): protection of women from these harmful trends is pivotal. What specific measures can be taken to protect female politicians/public figures.

Q by Alexandra Geese (Greens): there is also this issue of fake porn. I am very frustrated with the gender-based violence directive and the fact that serious harm has to be proved. This is a slap in the face for women and we have seen deepfakes/fake porn in every elections almost. Once you get pictures out, you cannot get them out of people's brains. What can the COM do about this. Did you believe watermarking really helps?

Q by Marc Faddoul: when you make the problem serious enough for platforms to deal with, such as with copyright and porn content, so if you have a big enough hammer, they will create the means to tackle it. Do we need a bigger hammer here?

Q by Tanya O'Carroll: it is staggering to see how much it turns women away from politics. We have been hammering on the same recommendations to the platforms for years. Very little is done, and I think the DSA will be a game changer in this. We cannot only deal with this on a content layer. There is a whole bunch of abuse that will be blocked, but it will probably be kept in a corner of the internet anyways.

- **Renate Nikolay**: About the VAW Directive, you don't have to convince me. It is shocking how little result it had. VAW is amongst systemic risks of the DSA. We have to start somewhere. We have not forgotten that VAW is one of the systemic risks that has to be mitigated. All the things in the guidelines can help. It will probably not avoid all deepfakes and all disinformation.

Q by Katalin Cseh (Renew): As someone from Hungary, I am very concerned about the independence of the national authorities that the DSA creates. Is this on your radar?

- **Renate Nikolay**: we have a role as guardian of the truth. Regarding the DSA coordinators, we have to ensure their independence. However, it felt better in this DSA report to be surrounded by 27 and it also felt better that citizens at least now have an address to go to, but we will watch this.

Alexandra Geese (Greens):

- I think in the next mandate we will have to look at the DMA and all the antitrust issues, and look in a comprehensive way to addictive design, advertising, etc. We will have to work together with the COM on this.
- There is a lot of frustration of women in the room on how little has been done to protect women online and offline, so maybe for the next priorities Let's make sure they have the same rights of public participation. At the end of this mandate we are almost empty handed.
- On recommender system, I can see we have 2 visions here:
  - One vision is to look at how we can mitigate the risks. For example, by creating guidelines.

HIGHLY CONFIDENTIAL GOOG-HJCFG-00001485

Confidential - Not For Public Release

- Another vision is to open up and be more brave and come together on what could be a new design for recommender system. What could they look like? The vision should be to not only mitigate the risk, but design a new sort of recommender system.

Panel 2: CSO meet enforcement authorities roundtable - the future of recommender systems

Speakers:

- Katarzyna Szymielewicz, Co-Founder and President of the Panoptykon Foundation
- Arturo Bejar, Whistleblower, former Facebook employee and consultant for Instagram
- Dr. Martin Degeling, Technical Expert "Auditing of Recommendation Systems" at the Stiftung Neue Verantwortung
- Felix Kröner, Policy Advisor at RESET.tech
- Emily Gorcenski, Data Scientist, Thoughtworks
- Anne Marie Pollock, Director of Policy, "Democracy and Fundamental Rights", Coimisiún na Meán
- Dr. Julia Marquier, Head of Unit "Net Neutrality, Platform Monitoring, Artificial Intelligence" at the German enforcement authority, Bundesnetzagentur
  Dr. Prabhat Agarwal, Head of Unit of the 'Digital Services and Platforms' Unit at the European Commission's DG CONNECT

Paul Tang :

- We need to act on the recommender systems, we need to do more. We have panellists first, and national authorities later. What should recommender systems look like in the future?

Katarzyna Szymielewicz :

- We need to work hand in hand with the Commission. We need to target VLOPs and recommender systems. Large platforms make choices when they design their services. They are designing these black boxes with business objectives. These platforms aren't good for us or for democracies, they are good for profits. We need a new recipe for recommender systems. The DSA fully enforced, we can get there. The DSA is making it possible.
- We shouldn't have behavioural profiling by default. It should optimise for values rather than engagement. We need to be prompting user choices, including opening up curation sources by third party services. We need positive friction to disrupt compulsive behaviour and trigger reflection, and we shouldn't have any addiction designs features.
- We need quality, reliability, relevance. We need signals of credibility from users. We need to prompt conscious user choices and opening content curation to third party services. Large platforms can do a lot to prompt user choice. We have designs choices that can make a difference. We need to add friction in sharing and resharing, and we need to avoid autoplay, infinite scroll, counts on social validation signals, notifications turned off by default.

Arturo Bejar :

- 30% of Instagram users saw false or misleading information last week. 25% witnessed hate. 1% of these people submitted a report. 1 to 2 reports are acted upon. 0,2% of people who witness this kind of content make a change.
- We are seeing content adaptation with adversarial AI. About 99% of disinformation is not violating any policy. A recommender system might see disgust as an engagement point, which will increase distribution. This is making engagement more important than everything else.
- People don't have tools to deal with unwanted content. People are experimenting harm on Instagram. We need to allow people to say they don't want to see certain things. People don't feel supported by platforms. We need to do 3 things: we need to change recommender systems to allow explicit content and to reject some unwanted content. It's about protecting people from what they see, not what they can post.

Martin Degeling :

- We have been investigating Tik Tok, but our findings relate also to other platforms. Recommender systems are all about optimising quantity over quality, it's about the time spent and ads. Transparency of recommendations is still very bad, and information is difficult to find or misleading. We need to be able to disable personalised feed. Users should be able to experiment, and researchers are also in the front line for all this.

Felix Kröner :

- We need to explore alternatives to recommender systems. We need to offer other forms of ranking, with feed resets for example. Algorithms can be used to promote and demote content. These demotions can take many different forms. We saw how Twitter demoted Russian media accounts before it was bought by Elon Musk. It was efficient and reduced engagement, but we saw that lifting these restrictions made the engagement with these accounts balloon. Platforms use the means of algorithms for the wrong ends. We haven't seen a lot of actions

Confidential - Not For Public Release

recently on national elections. Recommender systems reward controversial and harmful content, but done differently it could have a positive effect.

Emily Gorcenski :

- We can do a lot more to regularise the use of algorithms. Content moderation isn't only the way it is currently used. We focus on the content and not the actors. We need to stress the impacts of these platforms on the "real world". Many mass shootings in the US were committed by people who have been influenced by specific and small group of policy. These posts don't breach any rules but are used to fuel an atmosphere of hate. We need to minimise these impacts and the effect of these networks. We need to look at recommender systems beyond engagement – we need to include it because it's easy to measure, but values are equally important.

Paul Tang:

- We need to allow users to take back control, to have more agency about what's happening on their feed. These platforms need the attention, need the engagement. What would you say to the enforcers ?

Katarzyna Szymielewicz :

- Platforms won't change on their own. They will continue to make money and might make less in the future, but they will still make some. We need to force them to do changes.

Artujo Bejar:

- I've told Instagram and everyone about the effects. **Platforms won't change on their own. We need transparency of harm done to users. We have suffered harm and admitted harm**. If platforms were forced to share more data on victims of disinformation and misinformation, it will force them to act. We need platforms to act on the way users flag content and to act on it.

Martin Degeling :

- We need to look at each platforms, to look at detailed business models, and to act on it.

Felix Kröner :

- We need platforms to enforce the policies they already have, and we need to look beyond simple content moderation.

Emily Gorcenski :

- We don't have enough reliable data on content. We need transparency and more participation from the long tail of users that don't feel they have a voice at the moment. We need to make a more sustainable moderation overall.

Anne Marie Pollock :

- 13 VLOPs are based in Ireland so we work closely with the Commission in the enforcement of the DSA. The Board is key for that. We have an online safety framework with the DSA, a national content moderation regulation and a draft safety framework and expected to come into effect this year. We work very closely with many different partners. We need a whole of society approach.
- **On recommender systems, the DSA will be very important. We have additional measures on safety by design which are key, and recommender safety**. We need age-appropriate content, we need to avoid having children spend excess amount online. We propose platforms do safety assessments. Platforms should have safety as a primary concern. We need privacy assessment; we need platforms to know their risks.

Julia Marquier :

- This year is very important and election will have a strong impact. In Germany, we have legislation coming up later this year to have a DSC established. We need to have a strong view on recommender systems, we need to impose risks assessments and proper mitigation. Providers are free to choose their method as long as they are proportionate and these measures are working. We need to investigate and intervene when it's necessary.
- **We are seeing that large amount of misinformation are disseminated with bots in small amounts of time. We need adequate responses.** We need to look at alternative recommender systems and to see if it's a viable option. We are open to the inputs from civil society.

Prabhat Agarwal :

- We adopted a strong legislation in a small amount of time and it's important to recognise this effort. We didn't foresee most of the harm that are appearing these days. Deepfakes was added during the negotiations, as well as the emphasis on harms on minors. Recommender systems were an object we were regulated for the first time. We are now applying these rules and it's not an easy challenge. **Engagement based ranking, recommender systems, are valuable enforcement related evidence. We need and receive plenty of high quality data, and**

Confidential - Not For Public Release

it will allow for high enforcement quality. Before, we hoped for reputational damage on platforms, but we now have the law that we can apply.

- **We opened 2 investigations on X and TikTok, and we sent plenty of requests for information as we suspect issues from platforms. All requests have been replied in timely manner and have included relevant information. we are being transparent with what we do, we need to explain what we're doing. We have already included information manipulation on X, and we have looked into it for other platforms too. We need to work together, our objective is to see meaningful change.**
- We are issuing guidelines for clear incidents, but we aren't considering elections as similar incident to October 7th for example. On elections, usually the bottlenecks are on response mechanisms and being able to have teams to answer.

Question from the audience:

- Are we looking into how media companies are getting similar to social media ? Such as for example using A/B testing on making statements more inflammatory to gain exposure.

Julia Marquier :

- Not in a large extent but we are seeing how media companies are getting more on social media and using similar tactics, so we might in the future.

Prabhat Agarwal :

- It's a very narrow practice. The DSA does include a provision on media diversity on social media, it's included as a general rule within the DSA.

Maria Ressa:

- Digital news needs to do something different. I use to think social media was changing the world for the better. The EU has won the race of the turtles – it is needed, but there is so much more happening all around. We accepted a thesis of AI and social media, and generative AI. Journalists were held liable when we lied. New gatekeepers realised that if they spread lies faster they make more money. Lies spread faster than truth. Social media and GAI will not work to save democracy, it will work towards making it more addictive. If the reward system is predictable, it works, but if it's unpredictable, it's a pattern the way casino works. There is a cascading failure. A lie is rewarding online, because it creates engagement. We have to do things, and we aren't doing enough.
- On Rappler, we have given up on big tech. we rolled out a matrix chat app connected to our news agency. It's secure, decentralised, open source federated community. We need to create a shared reality for most people. We can't have democracies if everyone is in their own world. We need a level of debate that allows for conversations. We need a course correction. We have to act. I'm thankful to all the activists. We have to call BS when Big Tech ask us to do something about these systems.
- January 6th happened because of bulk invitations on Facebook, allowing for the spread of these theories.

Guénolé Carré (he/him)
Political Intelligence | Senior Consultant
EU Transparency Register ID - [redacted]

From: [redacted] <[redacted]@google.com>
Sent: 29 February 2024 11:57
To: Google <[redacted]@political-intelligence.com>
Subject: Fwd: Hearing with Frances Haugen & Maria Ressa 6/3, 15-18:00, EU Parliament, JAN 6Q2: PROTECTING THE 2024 ELECTIONS: FROM ALARM TO ACTION

Hi team,

Would you please follow this session and send us a readout?

Thanks,

[redacted]

Confidential & Not For Public Release

---------- Forwarded message ---------
From: **GEESE Alexandra** <[redacted]@europarl.europa.eu>
Date: Thu, Feb 29, 2024, 11:00
Subject: Hearing with Frances Haugen & Maria Ressa 6/3, 15-18:00, EU Parliament, JAN 6Q2: PROTECTING THE 2024 ELECTIONS: FROM ALARM TO ACTION
To:

PROTECTING THE 2024 ELECTIONS: FROM ALARM TO ACTION

HOW THE DSA CAN PROTECT DEMOCRACY BY MEANINGFULLY ADDRESSING THE VIRALITY OF DISINFORMATION AND HATE SPEECH

Societies across Europe are becoming more polarised, and widespread disinformation undermines trust in democratic institutions and electoral processes. Online platforms have spurred the proliferation of conspiracy theories, and disinformation, and actively encouraged toxic polarization as an anti-democratic instrument. With the widespread accessibility of large language models, this phenomenon is set to be supercharged. In light of the looming risks facing the 2024 European elections, rigorous enforcement of the Digital Services Act (DSA) is crucial – and with the integrity of European democracy at stake, the responsibility is monumental.

This event will explore electoral integrity in the online information space with a particular focus on engagement-based ranking and incentivising the production of divisive content. How can DSA implementation support the integrity of elections? What can platforms do to stop harmful amplification? How do less toxic recommender systems look like and what are the alternatives for content ranking on social media platforms? What can the European Commission and enforcement agencies do to make sure that elections take place in a safe online environment?

This Hearing will be co-chaired by MEPs Katalin Cseh / Renew Europe, Alexandra Geese / Greens/EFA and Paul Tang / Progressive Alliance of Socialists and Democrats. It will be livestreamed on YouTube and followed by some drinks and snacks.

Room JAN 6Q2

European Parliament, Brussels

Confidential - Not For Public Release

CLICK HERE TO REGISTER

WITH A KEYNOTE BY:

- Frances Haugen, Former Facebook Employee & Whistleblower; Advocate for Public Oversight of Social Media (remote)

…AND CONCLUDING REMARKS BY:

- Maria Ressa, CEO of Rappler, Journalist, Author, Freedom Advocate and 2021 Nobel Peace Prize Winner (in person)

1. Panel: recommender systems, disinformation, elections and the DSA

Speakers:

- Renate Nikolay, Directorate-General for Communications Networks, Content and Technology, EU Commission
- Marc Faddoul, transdisciplinary technologist, expert on recommender systems and algorithmic audit, Director and Co-Founder of AI-Forensics
- Tanya O'Carroll, independent advisor, strategist and leader focused on tech accountability, human rights and social justice

2. Panel: CSO meet enforcement authorities roundtable - the future of recommender systems

Speakers:

- Katarzyna Szymielewicz, Co-Founder and President of the Panoptykon Foundation
- Arturo Bejar, Whistleblower, former Facebook employee and consultant for Instagram
- Dr. Martin Degeling, Technical Expert "Auditing of Recommendation Systems" at the Stiftung Neue Verantwortung
- Felix Kröner, Policy Advisor at RESET.tech
- Emily Gorcenski, Data Scientist, Thoughtworks
- Anne Marie Pollock, Director of Policy, “Democracy and Fundamental Rights”, Coimisiún na Meán
- Dr. Julia Marquier, Head of Unit “Net neutrality, Platforms and Artificial Intelligence” at the German enforcement authority, Bundesnetzagentur
- Prabhat Argwal, Head of Unit of the 'Digital Services and Platforms' Unit at the European Commission's DG CONNECT

Confidential - Not For Public Release

Exhibit 245

EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR COMMUNICATIONS NETWORKS, CONTENT AND TECHNOLOGY

Platforms Policy and Enforcement
Digital Services

Brussels, 4 June 2024
CNECT.F.2/PA

Subject: Platform Election Readiness – DSA Roundtables 17/05/24 & 03/06/2024

Company Participants: Apple, Google, LinkedIn, Meta, Microsoft, Pinterest, Snapchat, Wikimedia, X, YouTube

Minutes:

- On 17 May and 3 June 2024, the European Commission's services responsible for the enforcement of the DSA held election readiness roundtables with providers of very large online platforms and very large online search engines (VLOPSEs). This followed the publication of the DSA election guidelines and relevant obligations for elections under the DSA. The second roundtable of 3 June was also attended by Digital Services Coordinators (DSC) and other national competent authorities.
- At the outset, the Commission underlined that it is actively monitoring compliance of providers of VLOPSEs with such obligations and may, where necessary, take enforcement actions related to the elections.
- The meetings focused on areas of importance for elections such as VLOPSEs' incidence response mechanisms, Generative AI risks, threat response frameworks and cooperation with relevant stakeholders such as relevant national authorities and civil society organisations. VLOPSEs were questioned on their actions to ensure readiness in relation to major incidents, such as the assassination attempt of Slovakia's Prime Minister, Robert Fico on 15 May 2024.
- The providers of VLOPSEs highlighted the importance of the election guidelines and the stress test in the form of a tabletop exercise, which took place on 24 April, as informing their work in the run up to elections. In particular, the tabletop exercise facilitated cooperation and exchange between relevant stakeholders, including DSCs, Civil Society Organisations and other providers of VLOPSEs.
- The providers of VLOPSEs presented the specific actions and changes that they have made in relation to election readiness and that they are implementing in compliance with the DSA election guidelines on their services.
- Providers of VLOPSEs announced actions such as avoiding or limiting election-related 'hallucinations' of GenAI systems and implementing new policies or terms and conditions in relation to manipulated media and proactively labelling GenAI content.
- Providers of VLOPSEs also announced adjustments to their political advertisement policies, anticipating the Regulation on Political Advertising.
- Providers of VLOPSEs have increased cooperation and escalation channels with relevant Member State authorities, civil society organisations and fact checking partnerships. For example, VLOPSEs have engaged in cooperation with EDMO

Commission européenne/Europese Commissie, 1049 Bruxelles/Brussel, BELGIQUE/BELGIË – Redacted
Office: L-51 04/P060 – Redacted
Redacted @ec.europa.eu

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

and EDMO's regional hubs to gain contextual information at the Member State level.

- Providers of VLOPSEs are engaging in enhanced monitoring of potential risks during the electoral period, including the risk of coordinated inauthentic behaviour, both in relation to their platforms and the platform's advertising systems.
- Providers of VLOPSEs are providing access to authoritative information on elections, such as initiatives linking to reliable information on how to vote in the EU Parliamentary elections.
- Providers of VLOPSEs that are signatories to the Code of Practice on Disinformation provided information on their efforts to respond to flags from other signatories under the Rapid Response System under the Code, while underlying that they found this system helpful.
- Providers of VLOPSEs also provided information on the measures that they are taking in relation to data access. As highlighted in the DSA election guidelines, reliable access to data from providers of VLOPSEs is pertinent especially in election periods.
- The providers of VLOPSEs stated that they are actively monitoring and taking action on disinformation narratives and coordinated inauthentic behaviour on their respective platforms, including coordinated campaigns such as Doppelganger. The providers of VLOPSEs are examining the tactics, techniques and procedures being employed by such actors.
- In relation to incident response, the Commission questioned providers of VLOPSEs on their readiness for potential elevated risks following incidents during electoral periods. The providers of VLOPSEs all stated that they were actively monitoring the potential risks that could result in the activation of incident response measures and that they had procedures in place to ensure quick escalation during the elections.
- The Commission highlighted that providers of VLOPSEs should be prepared in the run up to the elections and may receive further urgent requests with short deadlines from the Commission, where necessary. After the European elections the Commission will also organise a roundtable with the VLOPSEs to discuss and take stock of platforms' risk mitigation measures taken during the elections.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_017257

Exhibit 246

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]

Sent: 6/17/2024 3:16:23 PM

To: Redacted@disinformationindex.org]; Redacted demagog.org.pl]; Redacted Redacted fidu.it]; Redacted activefence.com]; Redacted @rsf.org]; Redacted arcom.fr]; Redacted Redacted @fidu.it]; Redacted @feedback.org]; Redacted @lessurligneurs.eu]; Redacted Redacted logically.co.uk]; Redacted (TikTok) [Redacted tiktok.com]; Redacted aiforensics.org] Redacted Redacted @legitimate.net]; Redacted maldita.es]; Redacted @tiktok.com] Redacted Redacted microsoft.com]; Redacted kreakom.dk] Redacted google.com]; Redacted Redacted lessurligneurs.eu]; Redacted newsguardtech.com]; Redacted disinformationindex.org]; Redacted Redacted tiktok.com]; Redacted @eaca.eu]; Redacted @google.com]; Redacted Redacted democracy-reporting.org] Redacted @doubleverify.com]; Redacted (GDI) Redacted @disinformationindex.org]; Redacted activefence.com]; Redacted @globsec.org] Redacted eui.eu] Redacted funky.ong]; Redacted fidu.it]; Redacted doteurope.eu]; Redacted feedback.org] Redacted Redacted vimeo.com]; Redacted abeurope.eu]; Redacted wfanet.org]; Redacted Redacted @alliance4europe.eu] Redacted @legitimate.net]; Redacted newsguardtech.com] Redacted Redacted linkedin.com]; Redacted eeas.europa.eu Redacted doteurope.eu Redacted valid.tech] Redacted Redacted democracy-reporting.org] Redacted crispthinking.com] Redacted Redacted vosteurope.org] Redacted abeurope.eu] Redacted newtral.es] Redacted Redacted newsguardtech.com] Redacted eui.eu Redacted vosteurope.org] Redacted (CMS/ERGA) Redacted rpms.sk]; Redacted The Bright Ap Redacted thebrightapp.com] Redacted democracy-reporting.org] Redacted Redacted globsec.org] Redacted faktograf.hr]; Redacted valid.tech]; Redacted vimeo.com]; Redacted (Ebiquity) [Redacted ebiquity.com] Redacted vosteurope.org]; Redacted @google.com] Redacted Redacted vosteurope.org]; Redacted @globsec.org]; Redacted aiforensics.org] Redacted Redacted demagog.org.pl]; Redacted @eeas.europa.eu Redacted @fb.com] Redacted @rsf.org] Redacted Redacted avaaz.org]; Redacted @eui.eu]; Redacted @vimeo.com] Redacted @eui.eu]; Redacted logicallyfacts.com]; Redacted avaaz.org]; Redacted firma.seznam.cz] Redacted Redacted google.com]; Redacted aiforensics.org Redacted afp.com]; Redacted Redacted newtral.es]; Redacted maldita.es]; Redacted fb.com] Redacted Redacted ebiquity.com] Redacted @twitch.tv]; Redacted microsoft.com] Redacted Redacted @linkedin.com]; Redacted democracy-reporting.org]; Redacted ch-c.eu]; Redacted google.com]; Redacted eaca.eu]; Redacted avaaz.org]; Redacted justin.tv]; Redacted Redacted alliance4europe.eu]; Redacted firma.seznam.cz]; Redacted fb.com]; Redacted rsf.org]; Redacted eui.eu]; Redacted demagog.org.pl]; Redacted (NCRTV/ERGA) Redacted iag.gr]; Redacted firma.seznam.cz]; Redacted activefence.com]; Redacted crispthinking.com]; Redacted disinformationindex.org]; Redacted wfanet.org]; Redacted newsguardtech.com]; Redacted google.com] Redacted @aiforensics.org]; Redacted whotargets.me]; Redacted Redacted alliance4europe.eu]; Redacted logically.co.uk]; Redacted Active Fence Redacted activefence.com]; Redacted Redacted kreakom.dk]; Redacted @tiktok.com]; Redacted @logically.co.uk] Redacted Redacted @rvr.sk]; Redacted adobe.com]; Redacted @doubleverify.com]; Redacted @eeas.europa.eu; Redacted @eaca.eu] Redacted (ERGA) Redacted rvr.sk]; Redacted Redacted microsoft.com]; Redacted PagellaPolitica Redacted pagellapolitica.it]; Redacted abeurope.eu]; Redacted Redacted alliance4europe.eu] Redacted newsguardtech.com]; Redacted maldita.es] Redacted Redacted debunk.org] Redacted lessurligneurs.eu] Redacted @newsguardtech.com]

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

CC: [Redacted] logically.ai]; [Redacted] meta.com]; [Redacted] vimeo.com] [Redacted] [Redacted] @ec.europa.eu] [Redacted] ec.europa.eu]; [Redacted] meta.com]; [Redacted] [Redacted] ec.europa.eu]; [Redacted] arcom.fr] [Redacted] @doubleverify.com] [Redacted] [Redacted] ec.europa.eu]; [Redacted] ec.europa.eu]; [Redacted] ec.europa.eu]; [Redacted] [Redacted] @ec.europa.eu]; [Redacted] @ec.europa.eu]; [Redacted] @ec.europa.eu]; [Redacted] ec.europa.eu] [Redacted] ec.europa.eu]; [Redacted] @ec.europa.eu]; [Redacted] ec.europa.eu]; [Redacted] efcsn.com] [Redacted] @logically.ai]; [Redacted] [Redacted] democracy-reporting.org]; [Redacted] @ec.europa.eu] [Redacted] ec.europa.eu]; [Redacted] [Redacted] jabeurope.eu]

Subject: Seventh Plenary meeting of the Permanent Task-Force of the Code of Practice on Disinformation
Attachments: Agenda 7th Plenary of the CoP Task-Force.pdf
Location: https://ecconf.webex.com/ecconf/j.php?MTID=m9191b0334f72287c92a492bb8d004622

Start: 6/21/2024 1:00:00 PM
End: 6/21/2024 3:30:00 PM
Show Time As: Tentative

Recurrence: (none)

Dear all,

Please find attached the **final agenda** for the **Seventh Plenary of the Permanent Task-Force of the CoP on the 2024 EP elections**.

The meeting will take place on **21 June, from 15 to 17:30 CET on Webex**. Please find below the connection details.

We look forward to a rich discussion on the EP elections with you on Friday.

Best wishes,
The CoP team

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

DG CONNECT CNECT.I.4 is inviting you to a scheduled Webex meeting.

Need help? Go to https://help.webex.com

Friday, June 21, 2024

3:00 PM | (UTC+02:00) Brussels, Copenhagen, Madrid, Paris | 3 hrs

Join meeting

More ways to join:

Join from the meeting link

[redacted]

Join by meeting number

Meeting number (access code): [redacted]

Meeting password: [redacted]

Tap to join from a mobile device (attendees only)

[redacted] Belgium Toll

[redacted] Luxembourg Toll

Some mobile devices may ask attendees to enter a numeric password.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Dear all,

Please find attached to this email the **preliminary agenda** for our Seventh **Plenary meeting of the Permanent Task-Force of the CoP on the 2024 EP elections,** will take place **fully online** on **21 June.**
Please note that there may be minor changes to the schedule and speakers. We will send you the final agenda and the WebEx link closer to the event.

Best wishes,
The CoP Team

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 247

EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR COMMUNICATIONS NETWORKS, CONTENT AND TECHNOLOGY

Platforms Policy and Enforcement
Digital Services

Brussels, 21 June 2024

Dear Sir, / Dear Madam,

As emphasised in the guidelines for providers of Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) on the mitigation of systemic risks for electoral processes ([1]), safeguarding the integrity of the elections is a key priority for the European Commission.

Considering that the French snap parliamentary elections have been called on short notice, I would like to invite you to an online meeting on the subject, which will take place on **Monday, 24th of June 2024 from 11:00 hrs – 12:30 hrs CET**.

The agenda will guide the discussion as follows:

1. Welcoming and opening remarks	11:00-11:10
Redacted *Chairman of Arcom* Redacted *Director, Platforms Policy and Enforcement, DG Connect, European Commission*	
2. The French legislative elections – current events, legal and regulatory context	11:10-11:40
Presentation by Redacted *member of the Arcom Board* *Presentation by* Redacted *Head of Viginum (TBC)* *Presentation by* Redacted *Head of unit Digital services, DG Connect, European Commission*	
3. Roundtable and Q&A on specific electoral issues *Presentations from Google, Meta, Microsoft, TikTok and X on the French legislative elections, with a particular focus on the*	11:40-12:25

([1]) Commission Guidelines for providers of Very Large Online Platforms and Very Large Online Search Engines on the mitigation of systemic risks for electoral processes pursuant to Article 35(3) of Regulation (EU) 2022/2065 (C/2024/3014)

Commission européenne/Europese Commissie, 1049 Bruxelles/Brussel, BELGIQUE/BELGIË – Redacted
Office: L-51 04/DCS – Redacted

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

period of silence online, disinformation and political advertising followed by an exchange of views	
4. Closing remarks and end of the meeting	12:25-12:30

During the roundtable we will discuss your preparatory work for the French elections to be held on June 30 and July 7, including threat assessment and response, the main tactics, techniques and procedures you have detected, and response to incidents that might have occurred until now. Anything shared during the meeting shall remain confidential, unless agreed otherwise.

Please send the names of the participants from your service, and their functions within the company to this email address. Only those participants who have been registered will be accepted into the meeting to ensure confidentiality.

The Commission services further ask you to send an **acknowledgement of receipt** within 24 hours from the receipt of this letter to CNECT-DSA-Registry@ec.europa.eu.

The Commission services are available to answer any questions you may have related to this meeting.

Yours faithfully,

Redacted
Head of Unit

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 248

From: Redacted bytedance.com]
Sent: 6/24/2024 9:26:52 PM
To: Code Of Practice Task Force [CNECT-COP-TASK-FORCE@ec.europa.eu]
CC: Redacted meta.com]; Redacted google.com] Redacted Redacted vosteurope.org]; Redacted demagog.org.pl]; Redacted Redacted @ec.europa.eu]; Redacted arcom.fr]; Redacted Redacted feedback.org]; Redacted ec.europa.eu]; Redacted Redacted ec.europa.eu] Redacted @ec.europa.eu]; Redacted Redacted legitimate.net]; Redacted maldita.es]; Redacted tiktok.com]; Redacted microsoft.com]; Redacted newsguardtech.com]; Redacted Redacted bytedance.com]; Redacted google.com]; Redacted Redacted democracy-reporting.org]; Redacted (Active Fence) Redacted activefence.com]; Redacted Redacted ec.europa.eu]; Redacted globsec.org]; Redacted Redacted @legitimate.net]; Redacted ec.europa.eu]; Redacted Redacted linkedin.com]; Redacted democracy-reporting.org]; Redacted Redacted medienanstalt-nrw.de]; Redacted vosteurope.org]; Redacted Redacted pm.gov.pt] Redacted newtral.es] Redacted google.com]; Redacted globsec.org] Redacted faktograf.hr] Redacted Redacted vosteurope.org]; Redacted @globsec.org]; Redacted Redacted aiforensics.org]; Redacted ec.europa.eu]; Redacted Redacted logicallyfacts.com]; Redacted google.com]; Redacted google.com]; Redacted aiforensics.org] Redacted meta.com]; Redacted Redacted @twitch.tv]; Redacted microsoft.com] Redacted Redacted @linkedin.com] Redacted democracy-reporting.org]; Redacted Redacted alliance4europe.eu] Redacted demagog.org.pl]; Redacted (NCRTV) Redacted esr.gr]; Redacted @ec.europa.eu]; Redacted Active Fence) Redacted @activefence.com] Redacted agcom.it]; Redacted Google) Redacted google.com]; Redacted aiforensics.org]; Redacted @whotargets.me]; Redacted alliance4europe.eu] Redacted (Active Fence) Redacted activefence.com]; Redacted ec.europa.eu] Redacted eeas.europa.eu; Redacted Redacted rpms.sk]; Redacted microsoft.com] Redacted Redacted pagellapolitica.it]; Redacted bytedance.com] Redacted maldita.es]
Subject: Re: [External] Re: RRS French elections - mailing list

Hi All,

Thank you for reaching out. As indicated during the plenary, I can confirm on behalf of the TikTok team that we are happy to extend the RRS notification system to the upcoming French legislative election. Non-platform signatories of the code can report content to us using the following email: Redacted @tiktok.com.

Echoing what other platforms shared previously in this thread, it would be great to agree on the following principles:

- The RRS notification system is only applicable to the upcoming French election
- We kindly request the non-platform signatories to apply a high threshold to notify content (i.e. content that is time sensitive and poses serious or systemic concerns to the integrity of the French electoral process).
- The RRS notification system will be in place temporarily until a week after the vote (14 July 2024)

We are open to discussing further details during our weekly WG meeting tomorrow.

Thanks,
Redacted

From: Redacted google.com>

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Date: Mon, Jun 24, 2024, 19:17
Subject: [External] Re: RRS French elections - mailing list
To: Redacted @meta.com>
Cc: Redacted @vosteurope.org>, "Code Of Practice Task Force"<CNECT-COP-TASK-FORCE@ec.europa.eu>, Redacted demagog.org.pl>, Redacted Redacted @ec.europa.eu>, Redacted Redacted @arcom.fr>, Redacted @feedback.org> Redacted Redacted @ec.europa.eu>, Redacted @ec.europa.eu>, Redacted @ec.europa.eu>, Redacted legitimate.net>, Redacted Redacted maldita.es>, Redacted tiktok.com>, Redacted Redacted @microsoft.com>, Redacted newsguardtech.com>, Redacted Redacted bytedance.com>, Redacted google.com>, Redacted Redacted democracy-reporting.org>, Redacted Active Fence)" Redacted activefence.com>, Redacted @ec.europa.eu>, Redacted globsec.org>, Redacted legitimate.net>, Redacted @ec.europa.eu>, Redacted Redacted linkedin.com>, Redacted democracy-reporting.org>, Redacted medienanstalt-nrw.de>, Redacted Redacted vosteurope.org>, Redacted @pm.gov.pt>, Redacted Redacted newtral.es>, Redacted @google.com>, Redacted Redacted globsec.org>, Redacted faktograf.hr>, Redacted Redacted vosteurope.org>, Redacted globsec.org>, Redacted Redacted @aiforensics.org>, Redacted @ec.europa.eu>, Redacted logicallyfacts.com>, Redacted @bytedance.com>, Redacted google.com> Redacted google.com>, Redacted Redacted @aiforensics.org>, Redacted meta.com>, Redacted Redacted twitch.tv> Redacted @microsoft.com>, Redacted Redacted @linkedin.com>, Redacted democracy-reporting.org>, Redacted Redacted alliance4europe.eu>, Redacted demagog.org.pl>, Redacted Redacted NCRTV)"< Redacted esr.gr>, Redacted Redacted ec.europa.eu>, Redacted Active Fence)"< Redacted activefence.com>, Redacted agcom.it>, Redacted (Google)"< Redacted google.com>, Redacted Redacted aiforensics.org>, Redacted whotargets.me>, Redacted Redacted alliance4europe.eu>, Redacted (Active Fence)"< Redacted activefence.com>, Redacted @ec.europa.eu>, Redacted eeas.europa.eu"< Redacted eeas.europa.eu>, Redacted Redacted rpms.sk>, Redacted @microsoft.com> Redacted Redacted pagellapolitica.it>, Redacted bytedance.com>, Redacted Redacted maldita.es>

Dear all,

Thank you very much for reaching out - and thanks again for everyone's collaboration on drafting and applying the Rapid Response System (RRS) for the EP 2024 elections. We're proud to play a part in the process.

As indicated last week, we agree with using Google ([redacted]@google.com) and YouTube ([redacted]@google.com) aliases for the upcoming French elections, taking into account the following principles:

- The notifications can only apply to the upcoming French election. This is an exceptional extension of the RRS as originally conceived for the EP 2024 elections.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

- While we will do our best to review the notifications in a timely manner and provide high-level feedback on our actions, using the aliases doesn't equal the RRS that was negotiated specifically for the EP 2024 elections.
- As [redacted] suggests, we kindly ask the non-platform signatories to apply a high threshold to notify content (i.e. content that is time sensitive and poses serious or systemic concerns to the integrity of the French electoral process).
- This notification system will be in place temporarily - we can only maintain the alias until a week after the vote, 14 July 2024.

Please do not hesitate to reach out in case of any questions and thanks again!

Redacted on behalf of Google and YouTube teams

On Mon, Jun 24, 2024 at 7:13 PM Redacted meta.com> wrote:

Hi Redacted and all,

I should have explained the thinking behind that indeed. Our expectation is that content meeting the threshold of being time sensitive and posing serious or systemic concerns to the integrity of the electoral process taking place in France would most likely be in French given the nature of the election and campaign. If you do encounter content in another language which meets this threshold, you could include it in the form, provided you expressly state in the report the language of the content so that it can be routed appropriately.

Happy to discuss further tomorrow!

Thanks,

Redacted

Meta

Redacted

Public Policy Manager | EU Affairs

Meta | Mobile Redacted

From: Redacted vosteurope.org>

Sent: 24 June 2024 11:52

To: Redacted meta.com>

Cc: Code Of Practice Task Force <CNECT-COP-TASK-FORCE@ec.europa.eu>; Redacted <Redacted demagog.org.pl>; Redacted ec.europa.eu>; Redacted arcom.fr>; Redacted feedback.org>; Redacted Redacted @ec.europa.eu>; Redacted ec.europa.eu>; Redacted Redacted ec.europa.eu>; Redacted @legitimate.net>; Redacted Redacted maldita.es>; Redacted @tiktok.com>; Redacted Redacted microsoft.com>; Redacted newsguardtech.com>; Redacted Redacted @bytedance.com>; Redacted google.com>; Redacted <Redacted democracy-reporting.org>; Redacted (Active Fence) <Redacted activefence.com>; Redacted <Redacted @ec.europa.eu>; Redacted globsec.org>; Redacted <Redacted legitimate.net>; Redacted @ec.europa.eu>; Redacted <Redacted linkedin.com>; Redacted democracy-reporting.org>; Redacted <Redacted medienanstalt-nrw.de>; Redacted @vosteurope.org>; Redacted <Redacted pm.gov.pt>; Redacted newtral.es>; Redacted google.com>;

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted globsec.org>; Redacted faktograf.hr>; Redacted <Redacted vosteurope.org>; Redacted globsec.org>; Redacted Redacted@aiforensics.org>; Redacted ec.europa.eu>; Redacted Redacted@logicallyfacts.com>; Redacted bytedance.com>; Redacted Redacted google.com>; Redacted @google.com>; Redacted @aiforensics.org>; Redacted Redacted meta.com>; Redacted twitch.tv>; Redacted <Redacted microsoft.com>; Redacted @linkedin.com>; Redacted democracy-reporting.org>; Redacted google.com>; Redacted alliance4europe.eu>; Redacted Redacted demagog.org.pl>; Redacted (NCRTV) <Redacted esr.gr>; Redacted ec.europa.eu>; Redacted (Active Fence) <Redacted activefence.com>; Redacted agcom.it>; Redacted (Google Redacted @google.com>; Redacted Redacted@aiforensics.org>; Redacted whotargets.me>; Redacted alliance4europe.eu>; Redacted (Active Fence) Redacted@activefence.com>; Redacted @ec.europa.eu>; Redacted @eeas.europa.eu Redacted eeas.europa.eu>; Redacted @rpms.sk>; Redacted microsoft.com>; Redacted pagellapolitica.it>; Redacted Redacted bytedance.com>; Redacted maldita.es>

Subject: Re: RRS French elections - mailing list

This Message Is From an External Sender

Hello all,
Thank you @Code Of Practice Task Force team for kicking this off.
Redacted we fail to understand the limitation on French language only content.
"Moroccan immigrants form the second largest immigrant population in France – behind the Algerians and ahead of the Portuguese – with 836,400 people, according to INSEE data for 2022. Their descendants number around 1 million"

Due to the nature of these elections, and as we observed in other national elections, we expect that disinformation related content to be multilingual, and this limitation would only hinder those efforts.

Can this be taken into consideration from your side?
Best,
Redacted

On Mon, 24 Jun 2024, 10:22 Redacted meta.com> wrote:

Dear all,

Many thanks COP team for kicking off this thread. On the Meta side, as discussed last week, we would be grateful if we could agree to the following framing principles:

- Please continue to apply the same criteria/threshold to report content (i.e. content that is time sensitive and pose serious or systemic concerns to the integrity of the electoral process).
- Please just use the content form (not the email alias).
- The content flagged via the RRS for the French elections should all be French language content (and as before, please indicate that the language is French in the form when reporting the content).
- The email list below should indeed be updated today COB. While we can update email addresses to the extent that is needed, we ask not to add any brand new organisations. The RRS should continue to be for the organisations who have been part of the WG when the framework was put in place.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_016545

- Like the EU Election RRS, this will be in place temporarily. We can apply the same timing as previously and maintain it until a week after the vote, 14 July.

Many thanks,
Redacted

[redacted] (she/her)
Public Policy Manager | EU Affairs
Meta | Mobile Redacted

From: CNECT-COP-TASK-FORCE@ec.europa.eu <CNECT-COP-TASK-Redacted ec.europa.eu>
Sent: 24 June 2024 10:41
To: Redacted demagog.org.pl> Redacted Redacted @ec.europa.eu>; Redacted arcom.fr> Redacted <Redacted feedback.org> Redacted ec.europa.eu> Redacted <Redacted ec.europa.eu> Redacted ec.europa.eu> Redacted <Redacted legitimate.net> Redacted maldita.es> Redacted <Redacted tiktok.com> Redacted microsoft.com> Redacted <Redacted newsguardtech.com>; Redacted bytedance.com> Redacted <Redacted google.com>; Redacted democracy-reporting.org>; Redacted (Active Fence) <Redacted activefence.com>; Redacted @ec.europa.eu>; Redacted <Redacted globsec.org>; Redacted legitimate.net> Redacted <Redacted ec.europa.eu>; Redacted linkedin.com>; Redacted <Redacted democracy-reporting.org>; Redacted medienanstalt-nrw.de>; Redacted <Redacted vosteurope.org> Redacted pm.gov.pt>; Redacted newtral.es>; Redacted google.com>; Redacted globsec.org>; Redacted <Redacted faktograf.hr>; Redacted vosteurope.org>; Redacted <Redacted vosteurope.org>; Redacted globsec.org> Redacted <Redacted @aiforensics.org>; Redacted ec.europa.eu> Redacted <Redacted meta.com>; Redacted @logicallyfacts.com> Redacted <Redacted bytedance.com> Redacted google.com>; Redacted <Redacted google.com>; Redacted aiforensics.org> Redacted <Redacted meta.com> Redacted @twitch.tv>; Redacted <Redacted microsoft.com>; Redacted linkedin.com> Redacted democracy-reporting.org> Redacted google.com>; Redacted alliance4europe.eu> Redacted Redacted demagog.org.pl> Redacted (NCRTV) <Redacted esr.gr>; Redacted ec.europa.eu>; Redacted (Active Fence) <Redacted activefence.com>; Redacted agcom.it>; Redacted Google) <Redacted google.com> Redacted Redacted @aiforensics.org> Redacted whotargets.me>; Redacted <Redacted alliance4europe.eu> Redacted Active Fence) <Redacted activefence.com> Redacted Redacted @ec.europa.eu>; Redacted eeas.europa.eu <Redacted eeas.europa.eu> Redacted Redacted rpms.sk>; Redacted microsoft.com>; Redacted <Redacted pagellapolitica.it>; Redacted bytedance.com>; Redacted maldita.es>
Subject: RRS French elections - mailing list

This Message Is From an External Sender

Dear all,

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_016546

Following our discussion at the plenary meeting last Friday, please kindly confirm the contact points and the modalities of use of the RRS for the French elections on the Excel sheet below **by today 24 June cob**.

Mailinglist RRS.xlsx

Let us know if you have any difficulties in accessing the document so we can help you make the changes.

Best,

The CoP team

Confidentialidade: Esta mensagem (e eventuais ficheiros anexos) é destinada exclusivamente às pessoas nela indicadas e tem natureza confidencial. Se recebeu esta mensagem por engano, por favor, contacte o remetente e alimine a mensagem e ficheiros sem tomar conhecimento do respectivo conteúdo e sem reproduzi-la ou divulgá-la.
Confidentiality Warning: This e-mail message (and any attached files) is confidential and is intended solely for the use of the individual or entity to whom it is addressed. If you are not the intended recipient of this message please notify the sender and delete and destroy all copies immediately.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 249

Ref. Ares(2024)4780388 - 03/07/2024

EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR COMMUNICATIONS NETWORKS, CONTENT AND TECHNOLOGY

Platforms Policy and Enforcement
Digital Services

Brussels, 3 July 2024

Dear Sir, / Dear Madam,

As emphasised in the guidelines for providers of Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) on the mitigation of systemic risks for electoral processes ([1]), and in the letter of 13 May from Roberto Viola, safeguarding the integrity of the 2024 European Elections is a key priority for the European Commission.

Following the previous technical roundtables on election readiness, which took place on the 17th of May and 3rd of June 2024, I would like to invite you to a third online technical roundtable on the subject.

The technical roundtable will take place on the **10th of July 2024 from 16:30 hrs – 18:00 hrs CET**.

During this roundtable we will discuss the actions VLOPS and VLOSEs have taken with regard to the DSA Election Guidelines, specifically their impact and effectiveness during the European elections. We would like to hear your lessons learnt from these elections and future improvements you plan to employ for the various elections in Member States taking place this year. The agenda will guide the discussion as follows:

1. Opening Remarks
2. Impact and effectiveness of actions taken with regard to DSA Election Guidelines
3. Lessons learnt European elections and future improvements
4. Closing remarks

The Commission also recalls the recommendations in the DSA election guidelines on a post-election review. The Commission recommends that such internal reports should include an assessment of whether the internal performance metrics and any other assessment criteria were met before, during and after the elections, lessons learned and areas for improvement. The Commission also recommends that providers of VLOPs and VLOSEs publish a public version of such post-election review documents. With a view of these recommendations, we ask you **to share your post-election review** with us before the roundtable. Should this review be not ready yet, we ask you to inform us when you

([1]) Commission Guidelines for providers of Very Large Online Platforms and Very Large Online Search Engines on the mitigation of systemic risks for electoral processes pursuant to Article 35(3) of Regulation (EU) 2022/2065 (C/2024/3014)

Commission européenne/Europese Commissie, 1049 Bruxelles/Brussel, BELGIQUE/BELGIË – Tel. +32 22991111
Office: L-51 04/DCS – Tel. direct line +32 229-63817

CONFIDENTIAL TREATMENT REQUESTED - NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY
META-119HJC-0001115

plan to finalise it. Please also share a link to the **public version of your post-election review**, or if not yet available, the information when you expect it to be published.

Please use the e-mail address: CNECT-DSA-Registry@ec.europa.eu to send any written information before the roundtable. If specified, the information received shall be treated as DSA confidential information and thus will not be shared with anyone outside the Commission services. In addition, anything shared during the roundtable shall remain confidential, unless agreed otherwise.

Please also send the names of the participants from your service, and their functions within the company to this email address. Only those participants who have been registered will be accepted into the online meeting to ensure confidentiality. The Commission services further ask you to send an **acknowledgement of receipt** within 24 hours from the receipt of this letter to CNECT-DSA-Registry@ec.europa.eu.

The Commission services are available to answer any questions you may have related to this meeting.

Yours faithfully,

Prabhat AGARWAL
Head of Unit

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 250

Message

From: [redacted]@meta.com]
Sent: 7/10/2024 11:49:44 AM
To: [redacted]@meta.com]; [redacted]@meta.com]; [redacted]@meta.com];

- one led by DG Justice that will be available early next year with elections integrity and cooperation with the industry as one of the themes covered;
- another led by DG CNECT aimed prior to the summer recess on actions taken in relation to the DSA Election Guidelines with input from the DSCs - while noting that this will not be about individual call outs/assessment of platforms.

Kind regards,
[redacted]

[redacted]
Compliance Lead | EU/UK Integrity Compliance Office
Dublin, Ireland
+[redacted]
[redacted]@meta.com
Meta

NOTICE: This email (including any attachments) may contain information that is private, confidential, or protected by attorney-client or other privilege. Unless you are the intended recipient, you may not use, copy, or retransmit the email or its contents.

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 251

11th Meeting of the EU Support Hub for Internal Security and Border Management in Moldova on "Countering Foreign Information Manipulation and Interference" **Agenda** *Venue: Chisinau, Sfanta Vinere street, 7, the MIA Integrated Center of Law Enforcement*	
18th September	
Chişinău time 08.30 – 09:00	Registration and coffee
Opening session with press	
09:00 – 09:30	Speakers: • Redacted Minister of Internal Affairs of the Republic of Moldova • **High level representative,** Delegation of the European Union to the Republic of Moldova • **High level representative,** European Union Partnership Mission in the Republic of Moldova
09:30 – 09:50	Group photo
09:50 – 10:00	Coffee break
Closed working sessions with selected media presence	
Session 1 **Countering FIMI from a whole-of-society perspective**	
10:00 – 12:00 Coffee break (not fixed, coffee available throughout the session)	*Panel description: This panel will address counter-FIMI strategy from a wider perspective. It will focus on formats of cooperation between civil society and government agencies such as joint working formats or regular meetings to exchange views on FIMI. The panel should also particularly focus on problems and possible improvements that both the government and media and civil society face when dealing with FIMI such as low funding, low transparency, working in silos, threats or attempts to intimidate journalists etc. The panel will also highlight the needs of the government when cooperating with civil society and media to streamline countering of FIMI while preserving the freedom of the press and freedom of expression.*

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	Introductory remarks and moderation (15 minutes): • Redacted, Deputy Secretary General at the Ministry of Interior, Republic of Moldova • Redacted, Adviser, European Commission/DG Home **Moderator**: IPRE (Iulian Groza) and investigative journalist (Mariana Rata/TV8, Alina Radu/Zdg) – TBC **Speakers** (aprox. 10 minutes presentations): • **Representative from API, fact checking organisation** {discuss FIMI cases with a focus on social media and cooperation challenges with government} • Redacted **- Representative of Watchdog (Coalition of NGOs for info space integrity)** {discuss overall FIMI threat level on social media} – TBC • Redacted **- Moldovan media representative, Ziarul de Gardă** {discuss how would a coalition between media and state institutions would look like to counter FIMI. Needs, challenges, limits to ensure freedom of media} — TBC • **Representative from IPIS**, Moldovan NGO active in strategic communication {discuss challenges of civil society cooperation with state institutions when it comes to countering FIMI} • Redacted **Director of Detector, UA civil society representative** {discuss how would a coalition between civil society and state institutions would look like to counter FIMI. Needs, challenges, limits to ensure freedom of speech} TBC • **Representative from NOKTA portal,** Russian speaking media (Găgăuzia) {discuss challenges in reaching Russian speaking audiences} • Redacted **- International Media Support** {discuss how would the government be more involved in supporting the media coalition to counter FIMI while safeguarding editorial independence} — TBC • **Representative from EEAS data team** {discuss about FIMI-ISAC) – VTC • **Q&A session**
12:15 – 13:00	Lunch break
	Session 2 **Lessons learnt: the national counter FIMI frameworks in the EU and Moldova** (closed session)
13:00 – 15:30	*Panel description: This panel will highlight best practices and challenges in EU Member States and Moldova in tackling FIMI in an integrated manner, based on a recent case study, preferably on elections. The focus should be on describing legal*

DRAFT

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Coffee break (not fixed, coffee available throughout the session)	*instruments as well as whether Standard Operational Procedures were used and the problems encountered. The cases should also highlight whether FIMI incidents could be connected to other hybrid activities that support disinformation narratives, such as illegal money flows, violent protests. etc* Introductory remarks and moderation (15 minutes): • Redacted Prime Minister of the Republic of Moldova • Redacted Head of the Centre for Strategic Communication and Countering Disinformation of the Republic of Moldova • Redacted, Adviser, European Commission/DG Home **Moderator**: Moldovan MIA – TBC **Speakers** (max 10 minutes presentations): • Redacted Head of the Centre for Strategic Communication and Countering Disinformation of the Republic of Moldova {present national counter FIMI framework state of play} • Redacted strategic adviser on hybrid threats, EUPM {comments on progress made on FIMI framework in Moldova} • Redacted Security and Intelligence Service, Deputy Director, Moldova {present threat assessment of FIMI with a focus on social media} • **Representative from ANCRETI** {present study case on restrictive measures to counter FIMI focusing on cooperation with other agencies} • **Representative from Audiovisual Council** {present study case on restrictive measures to counter FIMI focusing on cooperation with other agencies} • **Representative of the Ministry of Internal Affairs** representative {study case on FIMI supported by other hybrid instruments such as illegal migration, smuggling or illegal money flows} • **Representative of the Lithuanian crisis Centre** {present study case of FIMI supported by other hybrid instruments and how it was countered using counter FIMI framework} • **Representative of Sweden Psychological Defense Agency Centre** {present study case of FIMI supported by other hybrid instruments and how it was countered using counter FIMI framework} • **Representative from Finland** {discuss about *targeting* study case} – TBC • **Representative from Frontex and/or Europol from Moldova** {present study case of FIMI supported by other hybrid instruments such as illegal migration, smuggling or illegal money flows} • **Q&A session**

DRAFT

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Next steps, conclusions and closing remarks	
15.30 – 16:30	Present list of conclusions of the first day of the event (needs, challenges, possible steps forward on building national whole-of-society approach to FIMI) Speakers: • Redacted Head of the Centre for Strategic Communication and Countering Disinformation of the Republic of Moldova • Redacted Deputy Secretary General at the Ministry of Interior, Moldova • Redacted Adviser, European Commission/DG Home • **EUPM** representative
19th September	
Session 3 **Addressing disinformation risks on online platforms through digital services regulation: The state of play in Moldova and the EU**	
08:30 – 09:00	Registration and coffee
09:00 – 10:00	*Panel description: This panel will highlight best practices and challenges in Moldova and EU Member States in tackling FIMI and risks for electoral processes stemming from disinformation through regulation of online services and through cooperation with social media platforms. The panel will particularly review the Digital Service Act, its implementation in the EU and at EU Member State level, and the regulatory framework and state of play of DSA alignment in Moldova.* Introductory remarks and moderation (15 minutes): • Redacted State Secretary, Ministry of Economic Development and Digitalization Representative of the Ministry of Economic Development • Redacted Deputy Director General, European Commission/DG Connect • Redacted Head of Unit, European Commission/DG Near **Moderator**: Representative of EUPM, DG Connect or StratCom Centre (TBC) **Speakers** (max. 7 minutes presentations): • Redacted Head of the Centre for Strategic Communication and Countering Disinformation of the Republic of Moldova {review conclusions of previous day on national FIMI threats on social media and current cooperation frameworks of Moldovan authorities}

DRAFT

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	• Redacted, Director, Ministry of Economic Development and Digitalization Representative of the Ministry of Economic Development {on the commitment to transpose DSA in the Republic of Moldova, challenges regarding the institutional framework, interaction and cooperation with very large online platforms, including from the perspective of addressing content that represents disinformation} • **Representative from ANCRETI** (TBC) {share experience in countering FIMI focusing on cooperation with other agencies and how the DSA may enable countering FIMI more effectively} • Redacted Policy Officer, European Commission/DG **Connect** {share an overview and update of DSA implementation in the EU, how it addresses risks to electoral integrity and civic discourse and how the EU supports Moldova in implementing the DSA} • **Redacted** Head of Strategy and Statistics, ANCOM {share experiences of implementing the DSA at the national level, in particular regarding risks to electoral integrity and civic discourse} • **Q&A session**
Session 4 **Effectively mitigating disinformation risks on online platforms: Lessons learned from the European elections and how they can be applied in Moldova**	
10:00 – 10:50	*Session description: This session will focus on lessons learnt from the European elections to understand how collective responses to FIMI and disinformation threats to electoral integrity on online platforms can be improved in Moldova in the context of the upcoming elections.* **Moderator:** Redacted, Team Leader, European Commission/DG Connect **Speakers:** - Redacted Head of Unit, European Commission/DG Connect {on the lessons learnt while applying the DSA and the Code of Practice during the European elections} (10 minutes) - Redacted, Board Member, ARCOM {on lessons learnt while applying the DSA during the European elections and French legislative elections} (10 minutes) - **Representatives of Google, Meta and TikTok (TBC)** {on their respective policies, tools and mechanisms applied around the European elections and what they intend to apply in Moldova around the October elections and referendum} (30 minutes)
10:50 – 11:20	Coffee break

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Session 5 **Threat assessment and election readiness exercise**	
11:20 – 13:00 (Coffee and snacks available throughout the session)	*Session description: In this session representatives of online platforms, Moldovan authorities, civil society organisations, fact-checkers, and media, EU Digital Services Coordinators, and the European Commission will analyse key online threats to the integrity of elections in Moldova and simulate collective responses to selected threat scenarios.* **Moderators:** Redacted, Team Leader, and Redacted, Policy Officer, European Commission/DG Connect - Part 1: Instructions to all participants (in plenary, 10 minutes) - Part 2: Joint threat analysis (in moderated breakout groups, 30 minutes) - Part 3: Simulation of collective responses to selected threat scenarios (in moderated breakout groups, 45 minutes) - Part 4: Summary and conclusions (in plenary, 15 minutes)
Next steps, conclusions and closing remarks	
13:00 – 13.20	Summarize and discuss conclusions of the second day of the event (needs, challenges, possible steps forward) Speakers: • Redacted, Head of the Centre for Strategic Communication and Countering Disinformation of the Republic of Moldova (TBC) • Redacted, Head of Unit, European Commission/DG Connect (TBC)

DRAFT

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 252

From: [Redacted]microsoft.com]
Sent: 9/24/2024 7:47:26 AM
To: [Redacted]ec.europa.eu]; [Redacted]ec.europa.eu]; [Redacted] [Redacted]meta.com]; [Redacted]google.com]; [Redacted]bytedance.com]; [Redacted]bytedance.com; [Redacted]linkedin.com]
CC: [Redacted]ec.europa.eu]; [Redacted]@ec.europa.eu]; [Redacted]ec.europa.eu]
Subject: [External] RE: Moldova Elections - language on temporary mechanism

Thanks [Redacted] Will flag internally and adding [Redacted] for LinkedIn.

From: [Redacted]ec.europa.eu>
Sent: Tuesday, 24 September 2024 09:37
To: [Redacted]ec.europa.eu>; [Redacted]meta.com>; [Redacted] <[Redacted]google.com>; [Redacted]bytedance.com>; [Redacted]bytedance.com; [Redacted] [Redacted]microsoft.com>
Cc: [Redacted]ec.europa.eu>; [Redacted]ec.europa.eu>; [Redacted]ec.europa.eu>
Subject: [EXTERNAL] RE: Moldova Elections - language on temporary mechanism

Dear all,

Following up on [Redacted] email and on her behalf, thank you again for your agreement to participate in the temporary escalation mechanism for the Moldovan elections. This is much appreciated.

In order to operationalize it, we would suggest the following.

- Limited timing: from 1 October to 27 October. The first week (1 Oct-7 Oct) would be used as a setting up process, when all actors can ensure everything is in place and flags can be made but no answer/action is expected.
- Three meetings (two during, one after the elections) between all actors involved to discuss the situation and give oral feedback. As discussed, we will keep the meetings very targeted and streamlined, but based on previous experience we consider them essential for this type of exchange, in particular to give oral feedback and clarifications, complementing written exchanges.
- Same threshold than agreed during the EP elections: "content, accounts or trends that are time sensitive and pose serious or systemic concerns to the integrity of the electoral process."
- Flags to contain:
    - name of platform(s) concerned
    - the relevant URL and other relevant information needed to identify the content (e.g. time stamp; or URL to violating content tied to an account) (not screenshots)
    - indication of language of the content
    - summary of the reason why it's being escalated

Following our conversations and hearing your concerns, we have put in place the following:

- All flags will be filtered/streamlined by BROD, to provide their expertise regarding quality/threshold to be met for the flags. You will be able to share the email addresses/form to which flags can be done directly with them. They will be your only interlocutor in sending the flags.
- We will share with you the names of all organisations involved.
- We will set up a call with factcheckers and CSOs involved to brief them on how the process works, and insist on the necessity to respect the threshold of "serious or systemic concerns to the integrity of the electoral process". Tommaso will participate to share his experience under the Code as well.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_014573

We would be happy to finalise the process on a call on Thursday 26/9 from 4.30-5.30pm. For the meeting to be productive, please share with us your feedback in advance.

Thank you again for your collaboration,

Kind regards,
Redacted

From: Redacted @ec.europa.eu>
Sent: Wednesday, September 18, 2024 11:01 PM
To: Redacted meta.com>; Redacted Redacted google.com>; Redacted <Redacted bytedance.com>; Redacted bytedance.com; Redacted microsoft.com
Cc: Redacted ec.europa.eu>; Redacted <Redacted ec.europa.eu>; Redacted ec.europa.eu>
Subject: Moldova Elections - language on temporary mechanism
Importance: High

Dear All,

Great thanks again for your engagement regarding Moldova, which is much appreciated!

I am very much looking forward to see you/your companies representatives in Chisinau tomorrow!

Following previous conversations, I'm very happy to hear that you've all accepted to set up a temporary escalation mechanism with BROD in the context of the Moldovan referendum/elections, building upon the experience of the RRS during the EP elections. We very much appreciate this cooperation and thank you for taking this forward. My team will be in touch to iron out all the practical details. I am sure that based on our conversations, we can shape it in way that works from all sides.

In order to give credit to this achievement, following bilateral conversations with you, and taking on board the sensitivities you raised, we propose the language below to communicate on this. From our perspective, while this language leaves no doubt that this initiative is an exceptional, self-standing one, as explained, from our perspective, it is essential to make reference to the Code as the source of inspiration, also related to the collaboration during the EP elections. Furthermore, to do justice to this additional effort, we consider it as key to make it clear that this is a special mechanism, not just being in touch on a loose basis.

As time is pressing, I propose to move forward with this language unless we hear back from you by tomorrow (Thursday) noon. If you have any concerns, to be operational in this time-sensitive context, I suggest that your colleagues that are currently in Chisinau reach out to me directly to discuss. I will be at the venue from the start of the day.

We are very much looking forward to this collaboration,

Kind regards

"Drawing inspiration from the multi-stakeholder cooperation under the Code of Practice on Disinformation, including during the EP elections, four major online platforms (Meta, Google, TikTok, and Microsoft) have agreed to set up a temporary escalation mechanism with the relevant EDMO hub (BROD) with fact-checking expertise in the Moldovan informational ecosystem to flag time-sensitive content that present threats to the integrity of the electoral process."

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 253

2024 European Parliament Elections **Confidential Report**
Executive Summary TikTok Technology Limited ('**TikTok**') produces this confidential report pursuant to the 'Commission Guidelines for providers of Very Large Online Platforms and Very Large Online Search Engines on the mitigation of systemic risks for electoral processes pursuant to Article 35(3) of Regulation (EU) 2022/2065' ('**Elections Guidelines**'). In accordance with the Elections Guidelines, TikTok has undertaken a post-election review ('**Review**') following the Elections for the European Parliament that took place across 27 Member States from 6 to 9 June 2024 ('**EU Elections**'). This report has been prepared by TikTok following the Review and is to be read alongside TikTok's Crisis Report on the 2024 EU Parliamentary Elections, to be published on 24 September 2024 as part of TikTok's commitments under the Code of Practice on Disinformation. TikTok has comprehensive and effective measures in place to anticipate, address and mitigate the risks associated with electoral processes, including the risks associated with election disinformation in the context of the EU Elections. Herein, '**Election Period**' refers to the period from 6 May 2024 to 9 June 2024.

Risk mitigation measures employed for the EU Elections	
Below, we describe the key risk mitigation measures that TikTok employed for the EU Elections, and our assessment of the effectiveness of these measures, as well as lessons learned and areas for improvement.	
Enforcing our policies	
(I) Moderation capabilities	**Summary.** We have over 6,000 people dedicated to moderating EU language content. As they usually do, our teams worked alongside technology to ensure that we were consistently enforcing our rules to detect and remove disinformation, covert influence operations, and other violative content and behaviour that can increase during an election period.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	Assessment of effectiveness, lessons learned and areas for improvement. • As highlighted by the European Digital Media Observatory ('**EDMO**')'s Final Report, there was an increase in EU-related disinformation in the months prior to the EU Elections and the amount of content TikTok removed for violating its Civic and Election Integrity policy increased in Q1 2024. These data points illustrate the benefits and effectiveness of TikTok's well-placed efforts dedicated to increasing and improving its moderation measures undertaken prior and during the Election Period. • For example, ahead of the EU Elections, to ensure consistency across our moderation teams and optimise preparations, TikTok updated its Community Guidelines by adding new policies and enforcement actions for unverified elections claims, election misconduct, distribution of hacked materials, and significantly misrepresented authoritative civic information. • Additionally, we successfully adopted a dedicated escalation process for misinformation involving fact-checking partners, which helped to ensure the rapid verification of content identified as potentially containing disinformation. When such content was detected by our tools that monitored risks across 27 EU countries, it was fact-checked and enforced where necessary. • TikTok's moderation and enforcement against violative content during the elections operated effectively. • TikTok will continue to be vigilant during future election periods and devote cross-functional resources on a priority basis to its processes for handling election risk. • TikTok cascades updated policy guidance as and when needed to respond to on- and off-platform activity.
(II) Mission Control Centre: internal XFN collaboration	**Summary.** Ahead of the EU Elections, we established a dedicated Mission Control Centre ('**MCC**') space in our Dublin office, bringing together in-person employees from multiple specialist teams within our safety department. Through the MCC, our teams were able to provide consistent and dedicated coverage of potential election-related issues in the run-up to, and during, the EU Elections.
	Assessment of effectiveness, lessons learned and areas for improvement. • Given the complexity of the EU Elections and the fact that they took place at a national level and required coordination of 27 different national elections across 27 EU Member States, our assessment is that the MCC effectively allowed the Trust and Safety working groups and internal partners to provide consistent and dedicated coverage of potential election-related issues prior to and during the elections.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	• Additionally, the work of the MCC was aided by our fact-checking partners, who leveraged their position as local experts to identify and escalate possible misinformation trends. • TikTok invited Coimisiún na Meán to a presentation on election integrity and to visit the MCC in Dublin. These opportunities provided valuable sources of input and alignment. • We continue working on developing the effectiveness of the MCC, including by scaling the feedback provided by our fact-checking partners globally.
(III) Countering misinformation	**Summary.** Our misinformation moderators receive enhanced training and tools to detect and remove misinformation and other violative content. We also have teams on the ground who partner with experts to ensure local context and nuance is reflected in our approach. In the four weeks leading up to and including the EU Elections, we removed over **2,600 pieces of content** for violating our civic and election integrity policies, and over **43,000 pieces of content** for violating our misinformation policies. We removed over **96%** of violative misinformation content before it was reported to us, and over **80%** before receiving a single view by users. Some of the misinformation narratives we observed included narratives around migration, climate change, security and defence and LGBTQ rights.
	Metrics for the Election Period • Number of video removals for violating Civic and Election Integrity policies: 2,650 • Number of video removals for violating Misinformation policies: 43,544 • % of proactive video removals for violating Civic and Election Integrity policies: 95.3% • % of proactive video removals for violating Misinformation policies: 96.9% • % of video removals for violating Civic and Election Integrity policies before receiving a single view by users: 89.7% • % of video removals for violating Misinformation policies before receiving a single view by users: 80.5% • % of video removals for violating Misinformation policies within 24h: 90.5% • % of video removals for violating Civic and Election Integrity policies within 24h: 96.6% • % of reactive video removals for violating Civic and Election Integrity policies: 4.7% • % of reactive video removals for violating Misinformation policies: 3.1%
	Assessment of effectiveness, lessons learned and areas for improvement.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	• As risks relating to elections tend to be contextual and highly localised, TikTok will continue to take account of these factors in planning and executing its risk mitigation strategies. • For example, the emergence and continuation of two prominent wars (i.e., the War in Ukraine and the Israel/Hamas Conflict) and the coinciding high number of elections in 2024, leading to more complex and nuanced political opinions being shared and new opportunities for bad actors to promote polarising narratives, has increased the complexity of identifying and moderating potential misinformation. • Ahead of the EU Elections, we observed consistent patterns of keyword co-occurrences and their association with certain policy violations, especially around misinformation and harmful content. This provided helpful indications of recurring themes and areas prone to violations during the Election Period. • Additionally, we updated our ad moderation models to, for example, detect relevant political symbols and flags relevant to the EU Elections.
(IV) Fact-checking	**Summary.** Our global fact-checking program is a critical part of our layered approach to detecting harmful misinformation in the context of elections. The core objective of the fact-checking program is to leverage the expertise of external fact-checking organisations to help assess the accuracy of potentially harmful claims that are difficult to verify. Across Europe, we worked with 12 local fact-checking partners and covered at least one official language of every Member State during the EU Elections. These partners supported our work through proactive surfacing and flagging of potential disinformation; verification of content; and contribution to a repository of fact-checked claims and in-app interventions. We also collaborated with fact-checking partners to create media literacy content and produce the Election Speaker Series.
	Assessment of effectiveness, lessons learned and areas for improvement. • We successfully adopted a dedicated escalation process for misinformation involving fact-checking partners, to ensure the verification of content identified as potentially containing disinformation. When such content was detected by our trend detection tools that monitored risks across 27 EU Member States, it was fact-checked and enforced where necessary. • Our fact-checking partners leveraged their position as local experts to identify and escalate potentially violative content.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	• We also involved fact-checking partners in the creation of localised media literacy content during the EU Elections, which successfully helped to ensure that messages were relevant to our community and encouraged more engagement. • We continue working with our fact-checking partners to effectively verify potential misinformation claims, scale their proactive feedback to identify potentially violative content and misinformation trends, and contribute to media and digital literacy campaigns, including with respect to elections.
(V) Deterring covert influence operations	**Summary.** Threat actors sometimes try to target online platforms during elections, and we have dedicated experts working to detect, disrupt, and stay ahead of deceptive behaviours. To further increase transparency, accountability, and cross-industry sharing we introduced, earlier this year, dedicated covert influence operations reports.
	Assessment of effectiveness, lessons learned and areas for improvement. • TikTok works to address risks related to CIOs that attempt to disrupt election processes. During the Election Period TikTok did not detect CIO networks focussed on the EU Elections. • During the H1 2024 we have not detected any covert influence operations (**'CIOs'**) focussed on the EU Elections. However, based on our covert influence operations transparency reporting between January and June 2024, we successfully disrupted 2 CIO networks that we assessed to have targeted EU Member States with broader narratives: - a network assessed to be targeting political discourse in Germany consisting of 46 accounts with 12,837 followers; and - a network assessed to be targeting political discourse in Sweden consisting of 35 accounts with 109,537 followers. We also continued to remove accounts from previously disclosed CIO networks when they try to reestablish a presence on the platform and our efforts there are disclosed monthly on our CIO transparency page.
(VI) Tackling misleading AI-generated content	**Summary.** In addition to requiring creators to label any realistic AI-generated content ('**AIGC**') and launching, in September 2023, an AI-generated content label to help people do this, ahead of the EU Elections we further strengthened our approach to tackling risks associated with AIGC: • TikTok updated its existing 'Synthetic and Manipulated Media' policy and renamed it the 'Edited Media and AI-Generated Content (AIGC)' policy, which became effective in May 2024. Through these updates, we

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	adopted more commonly used and easily understood language when referring to AIGC, and clarified our existing prohibitions on AIGC showing fake authoritative sources or crisis events, or falsely showing public figures in certain contexts including being bullied, making an endorsement, or being endorsed. • As AI evolves, we continue to invest in combating harmful AIGC by evolving our proactive detection models, consulting with experts, and partnering with peers on shared solutions. • TikTok joined the Content Authenticity Initiative and the Coalition for Content Provenance and Authenticity, and became the first video sharing platform to put Content Credentials into practice. We launched the ability to read Content Credentials that attach metadata to content, which we can use to instantly recognise and label AIGC. This helped us to expand auto-labelling to AIGC created on some other platforms. • TikTok also joined industry partners as a party to the Tech Accord to Combat Deceptive Use of AI in 2024 Elections to work together to safeguard communities against misleading or deceptive AIGC in the election year.
	Metrics for the Election Period • Number of video removals for violating Synthetic and Manipulated Media policy: 1,176 • % of proactive video removals for violating Synthetic and Manipulated Media policy: 96.8% • % of reactive video removals for violating Synthetic and Manipulated Media policy: 3.2%
	Assessment of effectiveness, lessons learned and areas for improvement. • AIGC tools are now widely available for anyone to create manipulated content. Previously, this required skilled creators and could involve high cost. A wider variety of people can now create AIGC and at faster speeds. Misuse of AIGC may aggravate election misinformation, for example by depicting public figures giving political endorsements. Additionally, AIGC may potentially make it easier for bad actors or threat actors to increase the quality and volume of disinformation and misinformation or automate malign influence operations. • In TikTok's preparations for the EU Elections, we identified a risk associated with potential lack of visibility on AIGC and dependency on self disclosure by those posting AIGC. We updated our existing policies regarding AIGC and renamed them the 'Edited Media and AI-Generated Content (AIGC)' policy that went into effect on 17 May 2024 and became the first platform to put Content Credentials into practice.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	• Additionally, we introduced new models which have helped to assist in the detection and moderation of AIGC uploaded to the platform, including: (i) a 'visual model' to detect whether a relevant image or video was created using a generative AI model; and (ii) an 'audio model' to detect whether audio was created using a generative AI model. • To further improve TikTok's capacity to detect undisclosed AIGC, including ahead of future elections, TikTok continues to explore updates to its AIGC detection models, including by partnering with third party partners.
(VII) Government, Politician, and Political Party Accounts (GPPPAs)	**Summary.** Many European leaders, ministers, and political parties are joining TikTok on a regular basis - many of them did ahead of the EU Elections, including approx. 30% of the sitting Members of European Parliament (MEPs) and the EU Parliament itself. These politicians and parties play an important role on our platform - we believe that verified accounts belonging to politicians and institutions provide the electorate with another route to access their representatives, and additional trusted voices in the shared fight against misinformation. Before the EU Elections we engaged with both sitting MEPs and candidates from across the EU 27 to ensure that those using TikTok had the necessary information to do so safely and authentically. We strongly recommended Government, Politician, and Political Party Accounts (GPPPAs) to get verified. Verified badges help users make informed choices about the accounts they choose to follow. It's also an easy way for notable figures to let users know they're seeing authentic content, and it helps to build trust among high-profile accounts and their followers. In addition, GPPPAs are banned from placing ads on TikTok and from using monetisation features. Solicitations for campaign fundraising by GPPPAs are not permitted either on the platform: that includes content like a video from a politician asking for donations, or a political party directing people to a donation page on their website.
	Assessment of effectiveness, lessons learned and areas for improvement. • During the EU Elections, we worked to proactively safeguard against potentially impersonating accounts, for example through internally labelling GPPPAs. • Additionally, we found the EU Elections provided an opportunity to further engage with political stakeholders, including political parties and individual candidates, to ensure that they had the necessary information to use TikTok safely and responsibly, and recommend that they verify their TikTok accounts. • These efforts helped improve GPPPAs verifications and enforcement of TikTok's policies, including enforcing the restrictions that applied to these accounts (in relation to political ads, campaign fundraising, monetisation

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	features). • TikTok plans to review its GPPPA Election Campaign Fundraising Restriction policy in order to improve operability and enforcement.
<u>Directing people to trusted sources</u>	
(I) Investing in media literacy	**Summary.** We invest in media literacy campaigns as a counter-misinformation strategy. In 2024, ahead of the EU Elections, we collaborated with fact-checking partners to launch localised media literacy campaigns for each of the 27 EU Member States. This localised approach helped to ensure that messages were relevant to our community and encouraged more engagement. From 28 March 2024 we launched an individual in-app Election Centre for each of the 27 EU Member States covering 24 official EU languages. Working with electoral commissions and civil society organisations, these Election Centres connected people with reliable voting information, including when, where, and how to vote; eligibility requirements for candidates; and, ultimately, the election results. All Election Centres contained a media literacy section that provided tips and resources on how to identify misinformation and improve media literacy skills. This section included videos from our trusted fact-checking partners and local media literacy bodies: Agence France-Presse (AFP), Deutsche Presse-Agentur (DPA), Demagog.pl, Demagog.cz, Facta, Faktograf, Logically Facts, Newtral, Poligrafo, Delfi.lt, The Journal, Nieuwscheckers, Funky Citizens, DigiQ, Ostro. We also worked closely with the European Parliament to incorporate their own media literacy videos into the Election Centres, with a 'follow' button to the Parliament's own TikTok account as a further link to authoritative election-related information. We directed people to the Election Centres through prompts on videos, LIVEs and searches related to elections, as well as through push notifications. These Election Centres were visited more than 7.5 million times, and search banners were viewed more than 63 million times.
	Metrics for the Election Period: • Number of visits to the Election Center Pages in the EU: 7.5M • Number of views of video notice tags leading to the Election Centre Pages in the EU: 4.7B • Number of views of search banners leading to the Election Centre Pages in the EU: 63M

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	Assessment of effectiveness, lessons learned and areas for improvement. • The above metrics demonstrate the effectiveness and wide reach of, and broad engagement with, our Election Centre and its availability to our users. • We continue to elevate authoritative sources, civic information and media literacy through in-app product features (e.g. the Election Centre) to connect people to trusted information and provide them with more context about the content they view. • TikTok will continue to further explore and test ways to bring context to users about the content they view through labels, especially in relation to unverified election claims.
<u>External engagement at the national and EU levels</u>	
(l) Rapid Response System: External collaboration with COPD Signatories	**Summary.** TikTok took an active participation in the set up of the COPD Rapid Response System ('**RRS**') launched in April 2024 to streamline the exchange of information among civil society organisations, fact-checking partners, and online platforms. The RRS was a time-bound dedicated framework for cooperation among COPD signatories (**signatories**) during the EU Elections, which allowed non-platform signatories to flag time-sensitive content, accounts or trends that may present threats to the integrity of the electoral process. TikTok received 5 notifications through the RRS before the EU Elections, which were rapidly actioned and resulted in banning or geo-blocking of accounts and content removals for violation of Community Guidelines.
	Assessment of effectiveness, lessons learned and areas for improvement. • The RRS performed well as part of the interim agreement between relevant COPD signatories for the EU Elections. • Specific actions and findings were discussed during the weekly calls held by EDMO where relevant COPD signatories and platforms that received reports shared information and reported on actions taken. • The RRS collaboration provided TikTok with direct input on potential election misinformation during the EU Elections. This additional information helped TikTok to successfully identify, for example, the extent of misinformation trends, leading to further investigations and additional proactive measures to enforce TikTok's Community Guidelines.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	• Although not part of the original agreement, TikTok and other signatories agreed to leverage the structure and learnings from the RRS in the EU Elections for the 2024 French snap legislative election, attesting to some of the RRS's benefits. • TikTok will continue discussing with relevant signatories developing an RRS mechanism, which will allow TikTok to further operationalise its processes for responding to RRS reports.
(II) Engagement with local experts	**Summary.** To further promote election integrity, and inform our approach to the EU Elections, we organised an Election Speaker Series with local and regional external experts, particularly from the fact-checking community, who shared their insights and market expertise with our internal teams. Our fact-checking partners and local media literacy bodies have also supported our launch of the 27 localised Election Centres that included media literacy videos from them.
	Assessment of effectiveness, lessons learned and areas for improvement. • TikTok endeavoured to tailor its approach to the EU Elections to region-specific issues and leverage local insights to enhance enforcement effectiveness. • For example, TikTok worked closely with fact-checking partners to verify election information and to identify potentially violative content. • Additionally, TikTok worked with fact-checking partners and local media literacy bodies to create media and digital literacy campaigns. This localised approach sought to ensure messages were relevant to our community, encouraged more engagement through tailored messages and supported our work on leaning in local. • TikTok will continue to iterate and and improve its policies and processes by engaging with civil society organisations and other relevant third parties to understand local contexts and to obtain authoritative information regarding the conduct of elections. • TikTok also will continue to receive reports from Community Partners and subject matter experts and to engage with its fact-checking partners to receive their local expertise and insights.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

(III) Engagement with national authorities and stakeholders	**Summary.** In order to prepare for the EU Elections, as well as in the context of local elections that took place in H1 2024, our local/regional Government Relations teams engaged with national authorities, including local Digital Services Coordinators (DSCs) in many EU Member States: - Kommunikationsbehörde Austria in Austria - Konkurrence- og Forbrugerstyrelsen in Denmark - Autorité de régulation de la communication audiovisuelle et numérique (ARCOM) in France - Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen (BNetzA) in Germany - Autorità per le Garanzie nelle Comunicazioni (AGCOM) in Italy - Coimisiún na Meán in Ireland - Autoritatea Națională pentru Administrare și Reglementare în Comunicații (ANCOM) in Romania - Rada pre mediálne služby (Slovak Council for Media Services) in Slovakia In accordance with the provisions of the Digital Services Act (DSA), in order to facilitate smooth and efficient communications, TikTok has designated a single point of contact for national authorities, including DSCs. We made sure this information is easily accessible by publishing it on our European Online Safety Hub. We also onboarded some of these national authorities on our Community Partner Channel which offers them an additional and direct avenue to report content to us for review.
	Assessment of effectiveness, lessons learned and areas for improvement. • As noted above, in addition to engagements with local DSCs, ahead of the EU Elections, TikTok invited Coimisiún na Meán to a presentation on election integrity and to visit the MCC in Dublin. These and similar opportunities provide valuable sources of input and alignment. • TikTok will continue to iterate and and improve its policies and processes by engaging with national election authorities to understand local contexts and to obtain authoritative information regarding the conduct of elections.
(IV) Engagement at the EU-level	
Engagement with the European Parliament and EU political parties	**Summary.** Our teams have also been closely engaging with the European Parliament since mid-2023 in order to discuss our preparations for the EU Elections and answer questions about our policies.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	Our teams continued to actively engage with all European political groups and political parties, to explain our approach to elections, to help with account verification when requested, and answer their questions about our policies. To assist these efforts, our teams developed specific resources to help explain our policies and tools to interested politicians, and developed a new guide to account verification. In H1 2024 our teams also organised dedicated briefings for political parties at national level. At EU level, our teams were in close contact with all EU political groups and parties, and were available in both Brussels and Strasbourg for bilateral meetings in the lead up to the EU Elections.
	Assessment of effectiveness, lessons learned and areas for improvement. • The EU Elections provided an opportunity for us to engage with political parties and candidates to ensure they had the necessary information to use TikTok safely and responsibly and allowed us the opportunity to explain the differences between TikTok and other platforms, including differences in our Community Guidelines with respect to political advertising and content enforcement. • As part of these engagements, we recommended that the political parties and candidates verify their TikTok accounts. TikTok continues to consider how we can effectively provide information or demonstrations to political parties and candidates interested in using TikTok. • In preparation for the EU Elections, we worked closely with the European Parliament, including by incorporating the European Parliament's own media literacy videos into the Election Centres, with a 'follow' button to the Parliament's own TikTok account as a further link to authoritative election-related information.
Engagement with the EU authorities	**Summary.** Ahead of the EU Elections, our teams engaged with the European Commission (EC), including regarding the Guidelines for VLOPs on the mitigation of systemic risks for electoral processes pursuant to the DSA. Our teams also participated in events conducted by the EC. In particular, they took part in a simulation exercise to test our election readiness in April 2024 that was attended by civil society organisations and a number of European authorities, dedicated EU Elections roundtables in May and June 2024, and a post-election roundtable in July 2024.
	Assessment of effectiveness, lessons learned and areas for improvement. • TikTok participated in constructive discussions prior to the EU Elections with the EC and other industry members to share improvements and learning on addressing risks during the Election Period. • The roundtables offered valuable opportunities to learn about actions peer industry members were taking to

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	address common risks faced during elections and to learn about the Commission's concerns and priorities.

Contributions of researchers, civil society organisations and fact-checking partners
Summary. TikTok regularly engages with a wide range of independent experts, researchers, civil society organisations and fact-checking organisations to identify, mitigate and address risks facing TikTok and its users, including those relevant to the EU Elections. During the EU Elections, we worked with our 12 local fact-checking partners across Europe to help assess the accuracy of potentially harmful claims that are difficult to verify. Our fact-checking partners also proactively surfaced and flagged potential misinformation, verified content, contributed to a repository of fact-checked claims and in-app interventions, and collaborated with us to create media literacy content and produce Election Speaker Series. Additionally, TikTok's existing Community Partner Channel ('**CPC**') was available to support risk identification from onboarded NGOs who act as trusted flaggers to report suspected harmful content, including election misinformation. We also took an active part in the RRS during the EU Elections, which involved relevant COPD signatories as well as civil society organisations to streamline the exchange of information for the EU Elections. TikTok received 5 notifications through the RRS before the EU Elections from Democracy Reporting International, Science Feedback and Alliance4Europe, which were rapidly actioned and resulted in banning or geo-blocking of accounts and content removals for violation of Community Guidelines.
Assessment of effectiveness, lessons learned and areas for improvement. • Involving fact-checking partners in the creation of localised media literacy content during the EU Elections helped to ensure that messages in the Elections Centres were relevant to our community and encouraged more engagement. • We continue working with our fact-checking partners to effectively verify potential misinformation claims, scale their proactive feedback to identify potentially violative content and misinformation trends, and contribute to media and digital literacy campaigns.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

- TikTok also continues to receive reports from community partners and subject matter experts with respect to fact-checking insights, and incorporate local insights (including from researchers, civil society organisations and fact-checking partners) in its strategy to enhance enforcement effectiveness.
- The RRS performed well as part of the interim agreement between relevant COPD signatories for the EU Elections and provided TikTok with direct input on potential election misinformation during the EU Elections. This additional information helped TikTok to identify, for example, the extent of misinformation trends, leading to further investigations and additional proactive measures to enforce TikTok's Community Guidelines.
- Although not part of the original agreement, TikTok and other signatories agreed to leverage the structure and learnings from the RRS in the EU Elections for the 2024 French snap legislative election, attesting to some of its benefits.
- TikTok continues to consider the role and effectiveness of the RRS, alongside TikTok's existing response mechanisms and tools.

Exhibit 254

EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR COMMUNICATIONS NETWORKS, CONTENT AND TECHNOLOGY

Media Policy
Media Convergence and Social Media

Eighth Plenary Meeting of the Code of Practice's Permanent Task-Force

1 October 2024, 10:00 – 17:00

Room 4D - Borschette, Rue Froissart 36, 1040 Brussels, Belgium
Webex link: [redacted]

From 09:00	*Arrival & security check*
9:30-10:00	*Welcome coffee*
10:00-10:05	**1. Introductory remarks** **Redacted** *Head of Unit, DG CNECT I.4.*
10:05-10:15	**2. Welcoming remarks** **Redacted** *Director for Media, DG CNECT*
10:15-10:30	**3. Remarks from the French Digital Services Coordinator** **Redacted** *ARCOM / member of the European Board for Digital Services*
10:30-11:45	**4. Next steps on the Code conversion** **Redacted** *Head of Unit, DG CNECT I.4.* *Head of Unit, DG CNECT F.4.*
11:45-12:00	*Coffee break*

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

12:00-13:00	**5. Thematic session #1 – Initiatives to protect the Elections from different angles** Moderator: Redacted **a) Transparency of political ads** *WhoTargetsMe,* Redacted **b) Media literacy** *Google/YouTube,* Redacted **c) Promoting authoritative information** *TikTok,* Redacted
13:00-14:15	*Lunch*
14:15-14:45	**6. Updates from Signatories and on the Taskforce activities** Moderator: Redacted **a)** New signatory: *Legitimate,* Redacted **b)** Update on the Transparency Centre: *VOST,* Redacted
14:45-15:45	**7. Thematic session #2 – Elections and Artificial Intelligence** Moderator: Redacted **a) AI Chatbots** *AIForensics,* Redacted **b) Fact-checking and AI** *Meta,* Redacted *& EFSCN,* Redacted **c) AI literacy** *Microsoft,* Redacted
15:45-16:00	*Coffee break*

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

16:00-16:45	**8. Update on the latest round of Structural Indicators** Moderator: **Redacted** *ERGA,* **Redacted** *TrustLab,* **Redacted**
16:45-17:00	**9. Next steps & AOB** **Redacted** *Deputy Head of Unit, DG CNECT*

- *Note that after the meeting we suggest an informal gathering nearby* -

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014834

Exhibit 255

From: Redacted ec.europa.eu]
Sent: 10/23/2024 9:57:56 AM
To: Redacted bytedance.com]
CC: Redacted @ec.europa.eu]; Redacted tiktok.com
Subject: RE: [External] MD Escalation Mechanism

Dear Redacted

Thank you very much for checking internally and fully understood for this afternoon.

Have a good flight,
Best,
Redacted

From: Redacted bytedance.com>
Sent: Wednesday, October 23, 2024 11:56 AM
To: Redacted @ec.europa.eu>
Cc: Redacted @ec.europa.eu> Redacted
<Redacted bytedance.com>
Subject: Re: [External] MD Escalation Mechanism

Hi Redacted

Thank you for your email.

I'm checking internally now on the extension of the RRS for MD and I will get back to you in the next 24h.

Also, for today's meeting I might join a bit later. Apologies but I'm on a plane right now and the departure has been severely delayed. In case it's gets worse, I will make sure to follow up asap with the group on our feedback on the reports received since last week.

Kind regards,
Redacted

From: Redacted ec.europa.eu>
Date: Wed, Oct 23, 2024, 08:18
Subject: [External] MD Escalation Mechanism
To: Redacted bytedance.com>
Cc: Redacted @ec.europa.eu>, Redacted bytedance.com>

Dear Redacted

Let me first thank you for the good cooperation under the temporary escalation mechanism for Moldovan elections. The constructive and quick feedback has been very much appreciated.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

As you know none of the candidate reached 50% in the first round which leads the election to another round on 5 November.

In view of the current disinformation landscape in the country, we would see relevant for the mechanism to stay in place until then.

This would mean in practice an extension of two weeks (the current iteration finishing on the 27th):

- Operational from 28/10 to 10/10
- Two meetings on a need basis

Please let us know if you agree with such extension or if you have any question,

Thank you,

Best,

Redacted

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 256

Message

From: [redacted]@meta.com]
Sent: 11/7/2024 5:15:44 AM
To: DSA Inbox Group [redacted]@meta.com]
CC: Meta DSA Compliance Office [redacted]@meta.com]; [redacted]@meta.com]; [redacted]@meta.com]; [redacted]@meta.com]; [redacted]@meta.com]; [redacted]@meta.com]; [redacted]@meta.com]; [redacted]@meta.com]; [redacted]@meta.com]; [redacted]@meta.com]; [redacted]@meta.com]; [redacted]@meta.com]; [redacted]@meta.com]; [redacted]@meta.com]
Subject: Fw: Save The Date: DSA Elections Roundtable
Attachments: 20241107_DSA Election Roundtable_Questions_LF.pdf

Please see details on Elections Roundtable below, including attached list of questions. Kindly let me know who will attend in the room on the day, and who will attend via Teams. There will be at least one representative from the CO in attendance. I would like to confirm attendance to CnaM by tomorrow at the latest.

Thanks
[redacted]

From: Flore Bouhey Dwan <[redacted]@cnam.ie>
Sent: Thursday, November 7, 2024 12:12 PM
To: [redacted] <[redacted]@meta.com>; [redacted]@meta.com>
Cc: Christopher Lynch <[redacted]@cnam.ie>; Mary Lamey <[redacted]@cnam.ie>; Meta DSA Compliance Office <[redacted]@meta.com>; [redacted]@meta.com>
Subject: RE: Save The Date: DSA Elections Roundtable

Produced to HJC

This Message Is From an External Sender

Dear [redacted] and [redacted],

Following my last save the date email for the DSA Election Roundtable, we can confirm that the event will take place on Wednesday, November 13th, from 10.30am sharp – 1.30pm in the President's Room in the RDS, Ballsbridge. A light lunch will be served. Due to room capacity, a maximum of two representatives may attend. If you wish to have additional representatives, a Microsoft Teams invitation will be circulated separately.

As the DSC of establishment, we are drawing together a diverse group of relevant stakeholders including representatives from the European Commission, Electoral Commission, Coimisiún na Meán, VLOPs and VLOSEs to discuss various aspects relating to elections. An agenda will be shared in advance of the event. There will be a Q&A session at the event from 12:30-1:30pm. Topics covered will include:

- Risk Assessment for Irish Elections
- Processes & Procedures
- Candidate Experience, Participation, and Safety
- Media Literacy
- Advertising
- Data Access
- Critical Incidents

A list of questions for platforms is attached. Please note that this list is non-exhaustive, and the nature of the roundtable discussion will likely give rise to additional questions.

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

If you have any queries in the meantime, please do not hesitate to contact me. Please confirm your attendance, along with the names of your representatives, by return email.

Kind regards,
Flore

Flore Bouhey-Dwan
Stiúrthóir | Director of Platform Supervision

T: +[redacted] | [redacted]@cnam.ie

Coimisiún na Meán

1 Áras Shíol Bhroinn, Bóthar Shíol Bhroinn,
Baile Átha Cliath 4, D04 NP20, Éire.
1 Shelbourne Buildings, Shelbourne Road,
Dublin 4, D04 NP20, Ireland.

www.cnam.ie

-
Tá an ríomhphost seo agus aon iatán a ghabhann leis rúnda agus is leis an duine sin amháin a bhfuil siad seolta chuige/chuici a bhaineann siad. Muna duitse an ríomhphost seo, ní ceart é a léamh ná a scaoileadh chuig aon tríú páirtí. Iarrtar ort teachtaireacht a sheoladh chuig an seoltóir nó chuig [redacted]@cnam.ie, agus an ríomhphost seo a scrios.

This email and any files transmitted with it are confidential and intended only for the use of the addressee. if you have received this email in error, please notify the sender or [redacted]@cnam.ie immediately and delete this email. If you are not the intended recipient, any distribution or copying of this email is strictly prohibited.
Déan machnamh ar an timpeallacht, le do thoil, roimh an ríomhphost seo a phriontáil.

Please consider the environment before printing this email.

Produced to HJC

From: [redacted] <[redacted]@meta.com>
Sent: Friday 25 October 2024 13:56
To: Flore Bouhey Dwan <[redacted]@cnam.ie>
Cc: Christopher Lynch <[redacted]@cnam.ie>; Mary Lamey <[redacted]@cnam.ie>; [redacted] <[redacted]@meta.com>
Subject: Re: Save The Date: DSA Elections Roundtable

Dear Flore

Thank you for the notification of the planned Election Roundtable on Wednesday, 13 November and confirm Meta's attendance. Can you kindly let us know if there is a restriction of the number of attendees.

Kind regards
[redacted]
EU/UK Integrity Compliance Office

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

From: Flore Bouhey Dwan <[redacted]@cnam.ie>
Sent: Friday, October 25, 2024 12:24 PM
To: [redacted] <[redacted]@meta.com>; Meta DSA Compliance Office [redacted]@meta.com>
Cc: [redacted] [redacted]@meta.com>; [redacted] <[redacted]@meta.com>; Christopher Lynch <[redacted]@cnam.ie>; Mary Lamey <[redacted]@cnam.ie>
Subject: Save The Date: DSA Elections Roundtable

Dear [redacted]

Please find attached a Save the Date letter for attendance to a DSA Elections Roundtable. This will take place on Wednesday, November 13*th*. An exact timing and location for the Election Roundtable will be confirmed in the coming weeks, but we ask that you save the date in your diary. We expect this event to last approximately 3 hours.

As the DSC of establishment, we are drawing together a diverse group to discuss readiness to deal with online election threats and to assess online risks and mitigation measures in place or planned to the electoral process in Ireland. A more detailed agenda will follow soon.

We would appreciate that you confirm your attendance for Wednesday, 13 November at your earliest possible convenience.

Kind regards,

Flore

Flore Bouhey-Dwan
Stiúrthóir | Director of Platform Supervision

T: +[redacted] | [redacted]@cnam.ie

Coimisiún na Meán

1 Áras Shíol Bhroinn, Bóthar Shíol Bhroinn,
Baile Átha Cliath 4, D04 NP20, Éire.
1 Shelbourne Buildings, Shelbourne Road,
Dublin 4, D04 NP20, Ireland.

www.cnam.ie

Tá an ríomhphost seo agus aon iatán a ghabhann leis rúnda agus is leis an duine sin amháin a bhfuil siad seolta chuige/chuici a bhaineann siad. Muna duitse an ríomhphost seo, ní ceart é a léamh ná a scaoileadh chuig aon tríú páirtí. Iarrtar ort teachtaireacht a sheoladh chuig an seoltóir nó chuig [redacted]@cnam.ie, agus an ríomhphost seo a scrios.

This email and any files transmitted with it are confidential and intended only for the use of the addressee. if you have received this email in error, please notify the sender or [redacted]@cnam.ie immediately and delete this email. If you are not the intended recipient, any distribution or copying of this email is strictly prohibited.
Déan machnamh ar an timpeallacht, le do thoil, roimh an ríomhphost seo a phriontáil.

Please consider the environment before printing this email.

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 257

Message

From: Flore Bouhey Dwan [redacted]@cnam.ie]
Sent: 11/18/2024 9:06:45 AM
To: [redacted] [redacted]@meta.com]
CC: Christopher Lynch [redacted]@cnam.ie]; Mary Lamey [redacted]@cnam.ie]; Meta DSA Compliance Office [redacted]@meta.com]; [redacted] [redacted]@meta.com]; [redacted] [redacted]@meta.com]
Subject: RE: Save The Date: DSA Elections Roundtable

Dear [redacted]

Hope you are keeping well.

Thanks to your colleagues for attending in person and remotely the Elections roundtable organised by Coimisiún na Meán on Wednesday 13th of November 2024. We appreciated the participation and engagement.

As the time for the questions was restrained, we would like to follow up on your media literacy initiative which are part of the guidance published by the European Commission on the 26th of March 2024.

"**Media literacy initiatives.** Best practice for providers of VLOPs and VLOSEs is to collaborate on, implement, invest and engage in media literacy initiatives and campaigns focussing on elections to foster critical thinking and improve users' skills in recognizing disinformation and manipulation techniques, also related to generative AI. This could be achieved by:

i. Collaborating with **local media literacy organisations**, as well as relevant associations, groups and networks, financially supporting, sharing and integrating election related initiatives and campaigns on the platform, including by developing joint initiatives. Local media literacy organisations represent a valuable resource in terms of knowledge of local contexts and target audiences. The Commission recommends the use of the network of EDMO and its hubs and the Commission's Expert Group on Media Literacy to find the relevant organisations at Member State level.

ii. Developing and applying **inoculation measures** that pre-emptively build resilience against possible and expected disinformation narratives and manipulation techniques by informing and preparing users. Such measures should take into account the specific local context where they are carried out and should be complemented by other measures providing reliable information to users. Inoculation measures can take different forms, including e.g., gamified interventions, such as participation in online games on the generation of disinformation which encourages a critical reflection on the tactics used to influence (25), video or other types of content (26), and where feasible, should be implemented in-app to foster ease of access.

iii. When designing media literacy campaigns, it is recommended that providers of VLOPs or VLOSEs take into account specific narratives as well as tactics, techniques and procedures (TTPs) that contribute to systemic risks that are likely to occur before, during and after an election, in line with the approach of adapting mitigation measures to the **relevant national context and target audiences**. "

1. Have you made use of the 2021 Media Literacy Toolbox?
2. What media literacy initiatives do you have in place?
3. How will you encourage candidates to report, when they may be reluctant and feel like it won't matter?

We would request a response by no later than 25th November 2024.

The proposed date is proportionate to the information being requested and the number of questions being asked. Please note that we may use the information provided by you to perform our functions and comply with our legal obligations. An Coimisiún is not making this request on foot of any statutory power to compel the provision of information and, accordingly, the information you provide in response to the queries is to be provided on a voluntary basis. Any information you provide may be relied upon by An Coimisiún in the exercise of its functions which may include sharing relevant information with the European Commission or other public bodies.

Should you have any questions, please do not hesitate to contact me.

Produced to HJC

Kind regards,
Flore

Flore Bouhey-Dwan
Stiúrthóir | Director of Platform Supervision

T: +[redacted] | [redacted]@cnam.ie

Coimisiún na Meán

1 Áras Shíol Bhroinn, Bóthar Shíol Bhroinn,
Baile Átha Cliath 4, D04 NP20, Éire.
1 Shelbourne Buildings, Shelbourne Road,
Dublin 4, D04 NP20, Ireland.

www.cnam.ie

Tá an ríomhphost seo agus aon iatán a ghabhann leis rúnda agus is leis an duine sin amháin a bhfuil siad seolta chuige/chuici a bhaineann siad. Muna duitse an ríomhphost seo, ní ceart é a léamh ná a scaoileadh chuig aon tríú páirtí. Iarrtar ort teachtaireacht a sheoladh chuig an seoltóir nó chuig [redacted]@cnam.ie, agus an ríomhphost seo a scrios.

This email and any files transmitted with it are confidential and intended only for the use of the addressee. if you have received this email in error, please notify the sender or [redacted]@cnam.ie immediately and delete this email. If you are not the intended recipient, any distribution or copying of this email is strictly prohibited.

Déan machnamh ar an timpeallacht, le do thoil, roimh an ríomhphost seo a phriontáil.

Please consider the environment before printing this email.

From: Flore Bouhey Dwan <[redacted]@cnam.ie>
Sent: Thursday 7 November 2024 12:13
To: [redacted] [redacted]@meta.com>; [redacted] <[redacted]@meta.com>
Cc: Christopher Lynch <[redacted]@cnam.ie>; Mary Lamey <[redacted]@cnam.ie>; Meta DSA Compliance Office [redacted]@meta.com>; [redacted] <[redacted]@meta.com>
Subject: RE: Save The Date: DSA Elections Roundtable

Dear [redacted] and [redacted]

Following my last save the date email for the DSA Election Roundtable, we can confirm that the event will take place on Wednesday, November 13th, from 10.30am sharp – 1.30pm in the President's Room in the RDS, Ballsbridge. A light lunch will be served. Due to room capacity, a maximum of two representatives may attend. If you wish to have additional representatives, a Microsoft Teams invitation will be circulated separately.

Produced to HJC

As the DSC of establishment, we are drawing together a diverse group of relevant stakeholders including representatives from the European Commission, Electoral Commission, Coimisiún na Meán, VLOPs and VLOSEs to discuss various aspects relating to elections. An agenda will be shared in advance of the event. There will be a Q&A session at the event from 12:30-1:30pm. Topics covered will include:

- Risk Assessment for Irish Elections
- Processes & Procedures
- Candidate Experience, Participation, and Safety
- Media Literacy
- Advertising
- Data Access
- Critical Incidents

A list of questions for platforms is attached. Please note that this list is non-exhaustive, and the nature of the roundtable discussion will likely give rise to additional questions.
If you have any queries in the meantime, please do not hesitate to contact me. Please confirm your attendance, along with the names of your representatives, by return email.

Kind regards,
Flore

Flore Bouhey-Dwan
Stiúrthóir | Director of Platform Supervision
T: +[redacted] | [redacted]@cnam.ie

Coimisiún na Meán

1 Áras Shíol Bhroinn, Bóthar Shíol Bhroinn,
Baile Átha Cliath 4, D04 NP20, Éire.
1 Shelbourne Buildings, Shelbourne Road,
Dublin 4, D04 NP20, Ireland.

www.cnam.ie

-
Tá an ríomhphost seo agus aon iatán a ghabhann leis rúnda agus is leis an duine sin amháin a bhfuil siad seolta chuige/chuici a bhaineann siad. Muna duitse an ríomhphost seo, ní ceart é a léamh ná a scaoileadh chuig aon tríú páirtí. Iarrtar ort teachtaireacht a sheoladh chuig an seoltóir nó chuig [redacted]@cnam.ie, agus an ríomhphost seo a scrios.

This email and any files transmitted with it are confidential and intended only for the use of the addressee. if you have received this email in error, please notify the sender or [redacted]@cnam.ie immediately and delete this email. If you are not the intended recipient, any distribution or copying of this email is strictly prohibited.

Déan machnamh ar an timpeallacht, le do thoil, roimh an ríomhphost seo a phriontáil.

Produced to HJC

Please consider the environment before printing this email.

From: [redacted] <[redacted]@meta.com>
Sent: Friday 25 October 2024 13:56
To: Flore Bouhey Dwan [redacted]@cnam.ie>
Cc: Christopher Lynch <[redacted]@cnam.ie>; Mary Lamey <[redacted]@cnam.ie>; [redacted]
<[redacted]@meta.com>
Subject: Re: Save The Date: DSA Elections Roundtable

Dear Flore

Thank you for the notification of the planned Election Roundtable on Wednesday, 13 November and confirm Meta's attendance. Can you kindly let us know if there is a restriction of the number of attendees.

Kind regards
[redacted]
EU/UK Integrity Compliance Office

From: Flore Bouhey Dwan <[redacted]@cnam.ie>
Sent: Friday, October 25, 2024 12:24 PM
To: [redacted] <[redacted]@meta.com>; Meta DSA Compliance Office [redacted]@meta.com>
Cc: [redacted] <[redacted]@meta.com>; [redacted] <[redacted]@meta.com>; Christopher Lynch
[redacted]@cnam.ie>; Mary Lamey <[redacted]@cnam.ie>
Subject: Save The Date: DSA Elections Roundtable

Dear [redacted]

Please find attached a Save the Date letter for attendance to a DSA Elections Roundtable. This will take place on Wednesday, November 13*th*. An exact timing and location for the Election Roundtable will be confirmed in the coming weeks, but we ask that you save the date in your diary. We expect this event to last approximately 3 hours.

As the DSC of establishment, we are drawing together a diverse group to discuss readiness to deal with online election threats and to assess online risks and mitigation measures in place or planned to the electoral process in Ireland. A more detailed agenda will follow soon.

We would appreciate that you confirm your attendance for Wednesday, 13 November at your earliest possible convenience.

Kind regards,

Flore

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Flore Bouhey-Dwan
Stiúrthóir | Director of Platform Supervision

T: +[redacted] | [redacted]@cnam.ie

Coimisiún na Meán

1 Áras Shíol Bhroinn, Bóthar Shíol Bhroinn,
Baile Átha Cliath 4, D04 NP20, Éire.
1 Shelbourne Buildings, Shelbourne Road,
Dublin 4, D04 NP20, Ireland.

www.cnam.ie

-

Tá an ríomhphost seo agus aon iatán a ghabhann leis rúnda agus is leis an duine sin amháin a bhfuil siad seolta chuige/chuici a bhaineann siad. Muna duitse an ríomhphost seo, ní ceart é a léamh ná a scaoileadh chuig aon tríú páirtí. Iarrtar ort teachtaireacht a sheoladh chuig an seoltóir nó chuig [redacted]@cnam.ie, agus an ríomhphost seo a scrios.

This email and any files transmitted with it are confidential and intended only for the use of the addressee. if you have received this email in error, please notify the sender or [redacted]@cnam.ie immediately and delete this email. If you are not the intended recipient, any distribution or copying of this email is strictly prohibited.

Déan machnamh ar an timpeallacht, le do thoil, roimh an ríomhphost seo a phriontáil.

Please consider the environment before printing this email.

Produced to HJC

Exhibit 258

EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR COMMUNICATIONS NETWORKS, CONTENT AND TECHNOLOGY

Platforms Policy and Enforcement
Digital Services

Brussels, 28 November 2024

Subject: Invitation to Election Roundtable on Romanian Elections- 29 November

Dear Sir, Madam,

In light of the Presidential and Parliamentary Elections in Romania, the European Commission is writing to invite you to attend a Digital Services Act Election Roundtable being held **online tomorrow, 29 November**. The roundtable will be chaired by Renate Nikolay, Deputy Director of DG CNECT.

Ensuring election readiness and safeguarding elections is a priority for the European Commission under the DSA. This has been clear from the European Commission's actions both at European level for the European Parliamentary elections and at Member State level for national elections. Given the importance of protecting our democracies, the European Commission has focused on issues relating to civic discourse and electoral processes, as is clear from the publication of the DSA Election Guidelines.

Considering the Romanian Presidential elections, with the first round held on 24 November and the second round being held on 8 December and the Parliamentary elections to be held on 1 December the European Commission invites you to participate in a DSA election roundtable. The DSA guidelines outline that cooperation with relevant stakeholders is a key element to ensure the effective use of instruments and obligations in place. This is why the roundtable will involve a range of stakeholders including Very Large Online Platforms, relevant Romanian Authorities and Regulators and stakeholders from Civil Society Organisations.

The roundtable will take place **tomorrow 29 November from 16:00-17:30** and will be held **fully online**. We would appreciate if you could please confirm in writing who will attend from your organisation to CNECT-DSA-Registry@ec.europa.eu.

Yours sincerely,

Prabhat AGARWAL
Head of Unit

Commission européenne/Europese Commissie, 1049 Bruxelles/Brussel, BELGIQUE/BELGIË – Tel. +32 22991111
Office: L-51 04/P060 – Tel. direct line +32 229-87153

Prabhat.AGARWAL@ec.europa.eu

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

META-119HJC-0001107

Exhibit 259

TikTok

Romania Elections
Platform Integrity Briefing

November 28th 2024

Sensitive and Confidential - For Intelligence Purposes Only

Agenda

- Introduction
- Covert Influence Operations
- Intelligence Update
  - Telegram
    - Situation
    - What we've found so far on TikTok
  - TikTok Live
    - Situation
    - What we've found so far on TikTok

Sensitive and Confidential - For Intelligence Purposes Only

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Introduction

- In this meeting we are going to give an update on the status of our current investigations and intelligence surrounding protecting the integrity of dialogue on TikTok in the run-up to, and during, the Romanian elections.

Sensitive and Confidential - For Intelligence Purposes Only

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Covert Influence Operations

September 2024

Origin: Romania
Target: Romania

The individuals behind this network created inauthentic accounts in order to spread misinformation and promote narratives critical of the current Romanian government. The network was found to direct users away from TikTok to news websites which we assess to be managed by the network for the furtherance of its strategic goal.

Number of Accounts: 22
Number of Followers: 299,692

Currently Under Review

1. **A potential network of 12 accounts. The individuals behind this network created inauthentic accounts in order to amplify narratives in Romanian favorable to a political candidate, attempting to manipulate Romanian elections discourse. The accounts within the network were found to amplify the same videos repeatedly**
2. **A potential network of 23 Accounts we assess to be affiliated media with Sputnik media, which have targeted Romanian and Moldovan audiences, following our large scale takedown of Sputnik media accounts in September 2024.**
3. **#EchilibruSiVerticalitate (Balance and Verticality). Suspicious activity around potentially undisclosed paid political promotion videos.**
4. **Telegram: Current situational report to be covered in the next slides.**

Sensitive and Confidential - For Intelligence Purposes Only

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013617

Intelligence Update - Telegram

Situation

A Telegram Channel has been identified in the Romanian media as a means of spreading mass promotional media on TikTok, Facebook, Instagram and Youtube for the candidate Călin Georgescu.

TG Channel: Redacted

Sensitive and Confidential - For Intelligence Purposes Only

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Intelligence Update - Telegram

Off-Platform

Telegram Channel: Redacted has 3.77K subscribers and instructs users to share videos and use specific hashtags and comments on Facebook, Instagram, Youtube and TikTok.

What we've found so far on TikTok

Signal 1:
Users are encouraged to use #CG11 in all posts.
What we found:

- 1183 accounts had used this hashtag on video posts on TikTok.
- 157 of these accounts were registered on TikTok after the Telegram Channel creation date (15th June 2024).
- **76** of these accounts are assessed to be sole purpose accounts, created with the intention to post pro-Georgescu content. The remaining accounts are assessed to be organic users in support of Georgescu.

Signal 2:
Users are encouraged to post comments using a specific string of emojis and Georgescu's name.
What we found:

- 83 accounts posting these comments (2912 comments in total).
- The official account of Călin Georgescu had posted 23% of these comments.
- **25** of these accounts are assessed to be sole purpose accounts, created with the intention to post pro-Georgescu content.

Sensitive and Confidential - For Intelligence Purposes Only

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_013619

Intelligence Update - TikTok Live

Situation

An account under the handle of Redacted sent TikTok live gifts totalling 381,000 USD in a month span. 24 Oct - 24 November. We are currently assessing whether this user has attempted to promote Georgescu in the Romanian elections through offering payment to other users.

Redacted has been KYC verified by TikTok on 26th December 2023 with A Romanian ID document. This account belongs to Redacted

Redacted

Sensitive and Confidential - For Intelligence Purposes Only

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013620

Intelligence Update - TikTok Live

What we've found so far on TikTok

- Redacted sent gifts to 131 accounts in total within this period, of which:
  - 8 accounts were found to display Calin Georgescu's picture or symbols.
  - 1 account which was found to display the AUR party pictures or symbols.
  - These 9 accounts were sent gifts totalling 132,000 USD in one month.

Redacted also sent gifts to 25 accounts that posted comments mentioning Calin Georgescu, mainly in October and November, including on the day and the day after the campaign.

Sensitive and Confidential - For Intelligence Purposes Only

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013621

Any Questions?

Sensitive and Confidential - For Intelligence Purposes Only

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 260

EXTERNAL INFORMATION SERVICE

NOTE

Analysis of national security risks generated by the actions of state and non-state cyber actors on IT&C infrastructures, support for the electoral process

Unclassified extract from Note no. 0654 of November 28, 2024

- Russia has a history of interfering in the electoral processes in other states. If in the ex-Soviet space Russia was and is quasi-present through hostile actions of influence, in the West, the involvement has become much more evident starting with 2016/ the US presidential election.

Russia's interest in such hostile actions directed towards the West has increased in intensity, with an increasingly complex modus operandi, a very large share of the actions being carried out in the online environment, since: a) it is difficult to identify the connection between the operations themselves and individuals; b) Russian services were forced to adapt and made the transition from classic influence actions (managing human resources through traditional methods) to online actions.

- This year, several elections took place in the West, including the European Parliament elections/June 2024, with Russia conducting, to varying degrees, influence actions. The following preliminary conclusions regarding Russian interference emerged:

- by order of Kremlin, detailed sociological research was conducted on the target states (trends in public opinion, political perspectives of certain parties/candidates, measures taken by the authorities against foreign electoral interference) and what is the response capacity;
- informational aggression, including propaganda, and the use of artificial intelligence for rapid content creation were emphasised;

In the Western online space, Russia has made constant efforts to access the widest possible audience categories, by expanding the online infrastructure (varied portfolio of dissemination channels, official vectors, local trainers and troll networks) and diversifying the techniques for spreading (pro)Russian narratives, with an emphasis on their distribution at the local level (e.g. coordination of messages, increasing the quantity and quality of multimedia content, including using : artificial intelligence/AI software; hijacking and creation of hashtags, mainly on the X network);

- Moscow's interest in increasing the chances of pro-Russian, far-right, anti-system candidates, "pacifists" and representatives of nationalist movements is maintained. In the Kremlin's view, the European far-right is receptive to Moscow and is on an upward trend in popularity.

Russia has flooded the information space with divisive narratives and in support of vectors (persons or political formations) with views close to the Kremlin (extremists, nationalists, populists, anti-system political figures, etc.).

Tactics, techniques and procedures: creation and consolidation of extensive networks of online platforms and channels/groups on social media, development of support/pro-Kremlin networks, which will subsequently act as vectors for promoting Russian narratives, with an emphasis on local opinion leaders with Eurosceptic visibility/visions; diversification of methods to circumvent Western measures; organization of cultural events dedicated to Russian values/policies; taking statements by European officials out of context; fuelling conspiracy theories; creating deepfake content to defame inconvenient candidates.

The scheme of dissemination/amplification of messages remained decentralized: local pro-Russian satellites, clones of famous websites, accounts/channels and troll networks on social media platforms;

Objectives: amplifying the population's fears regarding the deterioration of the security situation; undermining trust in authority by discrediting European leaders, officials/ruling parties and victimizing the opposition (the "totalitarian Europe" thesis); highlighting the negative impact of their policies/decisions on the socio-economic situation; strengthening the position of Eurosceptic and extremist leaders/formations; eroding political and social support for Ukraine; amplifying social discontent;

- **capitalizing on the passivity** of the authorities to create gaps in public opinion.

■ during elections organized in the Republic of Moldova, techniques of adapting information manipulation according to demographic groups (e.g. nationalists, religious communities, rural population, vulnerable citizens) were noted, with the help of viral videos and emotionally charged images used to provoke strong negative reactions. The narratives are alternated according to 'political developments, ensuring their relevance, and the messages transmitted on Telegram, Facebook, Instagram, TikTok and VKontakte are coordinated and uniform.

• Romania is perceived by decision-making centers in Moscow as an enemy state (,,unfriendly"), and according to our data, Moscow adopts a policy of active deterrence towards Romania.

In Russia's view, our country:

- "provokes and threatens" Russia's security through the NATO and American military potential it hosts;
- "wants to solve its economic problems at the expense of Russia" (the Treasury issue);
- represents a direct competitor in the Republic of Moldova.

Romania - along with other states on NATO's Eastern Flank - has become a priority for Russia's hostile actions, with the Kremlin having a growing interest in influencing (at least) the MOOCL and the agenda in Romanian society in an electoral context through:

- propaganda and disinformation (including by using emerging technologies in activities targeted at specific groups and communities - e.g. by aggregating publicly available data - e.g. political, economic and media consumption preferences - and implementing generative artificial intelligence modules through which they can transmit adapted propaganda messages, in real time'. at the individual level};
- supporting Eurosceptic candidates and fueling anti-system movements, including by

"involving them in protests to shape the public agenda;

- encouraging discontent/provoking emotional reactions at the level of population, so as to put pressure on the authorities to reduce/stop support for Ukraine.

We appreciate that Romania is a target for Russian aggressive hybrid actions, including cyber attacks and information leaks (hacks'and leaks) and sabotage.

It has to be noted that during this year, the political situation in Romania was also addressed in political talk shows in Russia - Russian journalists are launching the idea that pro-Russian forces in Romania could obtain over 30% in the parliamentary elections. '

• The analysis of the way in which the Russian propaganda apparatus targeted Romania in 2024 highlights an approach:

- indirect, through the membership in NATO and in connection with the support granted to Ukraine and the Republic of Moldova, references being triggered by developments on these issues/spaces;
- directly, through information operations in the context of the increase in "Romania's threats to Russia's security" in the form of expansion on the Eastern flank, in the violation of national airspace by Russian

drones and disinformation about the burning of warehouses on the national territory by so-called Ukrainian refugees.

The messages aimed to: (i) divide society on issues such as the control exercised by the US/NATO over Romania, the security threats generated by NATO membership and support for Kiev; (ii) discredit NATO and Romania's response capacity, amplifying the population's distrust in the national defense capacity; (iii) erode the population's support for Romania's foreign policy decisions; (iv) highlight Romania's involvement in the conflict and territorial ambitions in relation to neighboring states (e.g. Ukraine and the Republic of Moldova).

In the case of information operations that directly targeted Romania, the modus operandi was similar to other propaganda actions launched in the European space since the beginning of the war in Ukraine: the same strategy of creating and validating fake news (through multimedia content and extensive false details), the use of similar methods of promoting and circulating the content (same initial source; same accounts involved in promoting the message; same audiences) and the erroneous connection of events with Ukrainian vectors (refuees or the resistance movement).

Tactics, Techniques and Procedures:

i) the use of photo-video materials whose veracity is difficult to prove - recordings of the alleged incidents (do not contain visible elements of identification, the initial source or location of the filming cannot be identified), accompanied by exaggerations regarding their magnitude; photos taken from older articles in the local press and manipulated (e.g. cropping the date of an image regarding a fire from July 31, 2024 in Bragadiru);

ii) mass rolling (cross-posting) through approximately the same accounts, on multiple social networks (X, Telegram, Facebook):. were engaged: Russian propaganda vectors aimed at the Russian public, including political analysts/ scientists or church-associated accounts, (pro)Russian accounts dedicated to covering the war in Ukraine and accounts with high visibility in the European space
(e.g. audiences in English, French, German, Turkish/Spanish), but also among the Arab and Chinese audience.

Such informational actions give credibility to subsequent news regarding sabotage incidents (true or false), generate panic and distrust in authorities and manipulate the collective mind (especially the audience with a low level of media culture) to place responsibility on the Ukrainian vector from the start.

Exhibit 261

From: Redacted tiktok.com]
Sent: 11/28/2024 2:37:59 PM
To: Redacted @ec.europa.eu]
CC: Redacted ancom.ro; Redacted roaep.ro; Redacted roaep.ro; Redacted ancom.ro; Redacted ancom.ro; Redacted roaep.ro; Redacted tiktok.com; CNECT-DSA-REGISTRY@ec.europa.eu; cnect-dsa-elections@ec.europa.eu; Redacted @ec.europa.eu]; Redacted tiktok.com]; Redacted @ec.europa.eu]
Subject: Re: [External] FW: TikTok follow up to roundtable on Romanian elections

Dear Redacted and team and representatives of the Romanian authorities

Thank you for the feedback received so far and we look forward to receiving any additional feedback the Romanian authorities may have to share.

In the meantime, we are glad to report that we are receiving illegal content take-down requests from the Romanian authorities and we are working our way through them. However, one point of order that we would ask to be adhered to in making these requests, is that we are provided with sufficient detail for us to be able to identify the alleged illegality, in a manner envisioned by Art. 9(2) of the DSA. At the moment, we have received a number of reports which are vague in nature or contain generic allegations of illegality without any specific laws or legal provisions being cited (e.g. the relevant law that has been said to be breached is not mentioned). It would also be helpful if content that is being reported as illegal is identified as precisely as possible. We have received one request which was very broad framed and essentially asked TikTok to monitor the platform for illegal content, which is obviously not in accordance with Art. 8 of the DSA.

Providing the requested detail, will help us in turning around requests as soon as possible (and allow us to do so in accordance with due process).

Thanks

Redacted

Redacted

Email: Redacted tiktok.com

On Thu, Nov 28, 2024 at 11:39 AM Redacted @ec.europa.eu> wrote:

> Dear President Redacted
>
> Dear Vice-President Redacted
>
> Dear Vice-President Redacted
>
> I would like to sincerely thank you and your teams again for participating in yesterday's ad hoc meeting on a first assessment of issues signalled in media and expert reports regarding the potential influence of content published and amplified on TikTok's social media platform on the preliminary outcome of Romania's Presidential Elections.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

As discussed yesterday, I herewith send you the topics and questions identified by TikTok. We have added one question that was not yet listed below in yellow.

We would be very grateful if you could review the issues and inform us whether you see issues that were discussed yesterday but have not been signalled yet, ideally **by today 16.00hrs**. TikTok can then provide answers which, in view of the on-going elections, should come at their earliest convenience; we would propose by **tomorrow 29 November cob at the latest**.

In addition, TikTok has provided a single point of contact for any issues regarding the elections as they proceed.

I would suggest that you perhaps also provide names of contact persons within your authority to whom TikTok representatives can reach out in case of questions or need for clarifications. TikTok may also want to reach out to you to establish higher level contacts in case urgent issues arise.

In addition, I would like to announce that tomorrow, the Commission is organising a Round Table bringing together platforms, the Romanian authorities and civil society organisations to discuss the Romanian elections. An invitation to participate will reach you separately.

Thank you very much for your attention paid to this issue and I propose that we stay in close contact,

Best regards,

Redacted

Redacted

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

European Commission

Directorate-General for Communication Networks Content and Technology

'Copernicus' building

L-51 – 4/028

Rue de la Loi 51
B-1000 Brussels/Belgium

Tel Redacted

Redacted ec.europa.eu

From: Redacted @tiktok.com>
Sent: Wednesday, November 27, 2024 8:21 PM
To: Redacted @ec.europa.eu>
Cc: CNECT DSA REGISTRY <CNECT-DSA-REGISTRY@ec.europa.eu>; CNECT DSA ELECTIONS <cnect-dsa-elections@ec.europa.eu>; Redacted @ec.europa.eu>; Redacted tiktok.com>
Subject: TikTok follow up to roundtable on Romanian elections

Dear Redacted and team,

Thank you very much for organising the call today to discuss the alleged issues relating to the elections in Romania. As we mentioned on the call, we treat election integrity and these matters with the utmost seriousness, especially in terms of the impact on democracy and role that the integrity of our platform plays in this regard.

We take this opportunity to address two points:

1. How to contact TikTok - to ensure clarity in this regard on the part of the authorities; and
2. Setting out the questions that we understand have been raised to date.

How to contact TikTok

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Taking on board the feedback received today and to ease communications, we would ask that the Romanian authorities use our DSA single point of contact for any issues as the election period proceeds: **DSA-single-point-of-contact-authorities@TikTok.com**

In addition, if the Electoral Authority would like to be onboarded to our TikTok Safety Enforcement Tool (TSET) platform, which will also allow it to flag content, we can arrange that. We look forward to hearing if this would be of interest.

Questions raised

As discussed at the meeting, set out below is a list of the questions that have been put to us by the European Commission and the Romanian authorities. If we have missed anything from the list, please do let us know. In particular, we appreciate that you had a number of questions on the Commission side and, as we did not have these in advance of the meeting, please do let us know if the list below does not record them in full.

We are also in the process of investigating the letter referenced by Mr Redacted from the Romanian Audiovisual Authority, which was stated to have been sent to TikTok. This was not a document that those of us on the call were aware of, and while we investigate if it has been received, it would be greatly appreciated if you could ask them (i) to confirm when and to whom it was sent and (ii) to provide a copy in the meantime in order to avoid any undue delays.

Questions / action items raised on the call

1. **Question for TikTok** - regarding the Romanian delegation that visited TikTok's Transparency & Accountability Centre in October 2024, TikTok is to set out who attended and what topics were covered.
2. **Questions for TikTok and the Romanian Authorities** - regarding the complaint sent to TikTok on 21 November 2024 and responded to on 22 November 2024 (the '**21 November Complaint**'):
    1. both TikTok and the Romanian authorities are to set out their approach to this complaint and the legal basis for their positions (plus clarify response times for this particular complaint); and
    2. in respect of the content that TikTok stated was geo-blocked in response to the 21 November Complaint, the Romanian authorities will provide the URLs to the content that they have said remains available in Romania notwithstanding TikTok's position that the content posted by politicians / political parties was geo-blocked.
3. **Question for the Romanian Authorities** - regarding responses more generally, the Romanian Authorities are to provide any communications that were sent and which they

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

consider were not responded to in a timely manner, so TikTok can look into these to see what happened and establish if any process changes are required.

4. **Questions from the Romanian authorities** - TikTok to explain how it engages with experts and local Member State authorities, including under the DSA, to address the assertion made that it is not paying due attention to relevant local authorities (e.g Ministry of Justice) who are actually responsible for administering the election and have local knowledge and understanding local legal requirements (e.g during 'silence periods').
5. **Question from the Commission** - seek an explanation from TikTok as to how to assess a situation where one candidate has had exponential growth in views.
    1. Related question from the Romanian Audiovisual Authority asking for responses from TikTok to a letter and enclosed document which sets out the questions it has regarding the increase in growth for the candidate in question. [**TikTok Note:** a*s flagged above, TikTok is looking for this letter but kindly asks that the Commission provide it in the meantime if they have it or are able to obtain a copy.*]
    2. Related Question from Romanian authorities: Please explain how TT's algorithm worked in the past two months in Romania in terms of amplifying content provided by the candidates or related to their discourse.

6. **Question from the Commission:** asking TikTok to set out its standard policy regarding the prohibition on political ads and whether the policy was effectively enforced in the Romanian election - in particular, have we identified such ads in Romania? [**TikTok Note:** as mentioned on the call, we do collect this data and it is typically produced, validated and shared after the relevant election].

7. **Question from the Commission:** regarding the potential role of paid influencers in the growth of this candidate, can TikTok set out its assessment of any such role (e.g to establish If they have been paid to promote content/the candidate)?

Thank you again, and kind regards,

Redacted

--

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted

Tel: Redacted
Email: Redacted @tiktok.com

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 262

Election Roundtable on Romanian Elections

29 November 2024

Questions to guide discussion

Election-Specific Risks and Tailored Mitigations

- Could you describe your internal processes for identification of specific risks for the RO Presidential and Parliamentary elections? At what point was this conducted?
- What particular risks have been identified?
- In relation to collecting and sharing relevant information, analysis, and data on RO specific risks to tailor mitigation measures, is there third-party cooperation in place, in general and specific to these elections? For example, cooperation with researchers, CSOs or fact-checkers
- Did you assess the capacity of all relevant mitigation measures to perform effectively in the RO linguistic and electoral context?

Resources for RO elections

- What resources are dedicated for the RO elections in particular? Do the teams working on these elections have specific linguistic and regional knowledge applicable to RO?

National electoral laws

- How do you make sure that national electoral laws are respected in RO?

Advertising

- Are there measures in place to identify ads with false candidate or party information?
- If an ad is promoted featuring false information how would that be addressed at content and account level?
- What resources are available to review political ads and advertisers when targeting Romania?

Data access

- What avenues do you have in place in relation to transparency and data access for researchers?

Critical incidents

- What processes and procedures do you have in place in relation to potential incidents and procedures during the electoral period?
- Which RO authorities have been onboarded to your escalation channels?

Media Literacy

- What media literacy initiatives do you have in place for the RO elections?

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 263

From: Redacted microsoft.com]
Sent: 11/29/2024 5:58:58 PM
To: Redacted meta.com]; Redacted ec.europa.eu] Redacted
Redacted @tiktok.com]; Redacted @linkedin.com]
CC: Redacted tiktok.com; Redacted tiktok.com] Redacted
Redacted @ec.europa.eu]; Redacted @ec.europa.eu] Redacted
Redacted ec.europa.eu]; Redacted google.com] Redacted meta.com] Redacted
Redacted meta.com]; Redacted meta.com]; Redacted
Redacted meta.com]; Redacted meta.com]
Subject: Re: [External] RO elections - Rapid Response System

Hi Redacted

From both the **Bing** and **LinkedIn** side (adding @ Redacted , we can also activate the RRS for Romania until December 12th and would agree with the points made below by Redacted in particular as to the scope.

The form for possible Bing related notifications can be accessed on the page https://aka.ms/EUCOP-Reports and they can in addition be sent by email to EUCOP-Reports@microsoft.com.

For LinkedIn, the email to send notifications to is also still the same: Redacted @linkedin.com

Best,
Redacted and Redacted

From: Redacted meta.com>
Sent: Friday, November 29, 2024 6:37 PM
To: Redacted ec.europa.eu>; Redacted tiktok.com>
Cc: Redacted tiktok.com <Redacted tiktok.com>; Redacted tiktok.com>;
Redacted @ec.europa.eu>; Redacted
<Redacted ec.europa.eu>; Redacted ec.europa.eu>; Redacted (Google)
<Redacted google.com>; Redacted microsoft.com>; Redacted
<Redacted meta.com>; Redacted meta.com>; Redacted
<Redacted meta.com>; Redacted meta.com>; Redacted
<Redacted meta.com>
Subject: Re: [External] RO elections - Rapid Response System

Dear Redacted and all,

Thank you for reaching out. Having aligned with my colleagues (a few of whom I have looped into this email thread) and following the roundtable just now, we can confirm we can participate in the RRS for Romania until December 12th.

Scope: We wanted to reiterate a couple of points on the parameters of the RRS which we've had in place in the past:

- The threshold for flags should be as before, content which poses a serious or systemic concern to the integrity of the election.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

- The scope should be limited to content in English or Romanian.
- We can acknowledge receipt of the flags but we do not commit to specific turn-around times. We will, as we have in the past, provide feedback where feasible and subject to limitations including confidentiality, user privacy, and abuse prevention.
- We do not commit to a specific content moderation outcome. We will review any content received independently.
- We ask that calls be held on a need-basis only (i.e. If there are flags and with those companies that received said flags), not more than once a week.

Operationalisation:

- **We will need an updated list of organisations and emails which will take part as soon as possible**. We think the system we had in place in the context of the election in Moldova (though outside the remit of the Code) where all flags went through an EDMO Hub/Funky Citizens would be preferable.
- Once we receive this list, we will put 2 things in place:
    - We will onboard the relevant organisations/emails to the same channel we have used for the European Parliament and the French snap election early next week.
    - In the meantime - and provided we have been given their details - should those organisations have flags over the weekend in view of the parliamentary election on Sunday, they can send them to my colleagues Redacted meta.com) and Redacted Redacted meta.com) with in the subject line "RO Election Rapid Alert System". **For this weekend especially, we would appreciate using a single organisation, such as the EDMO Hub/Funky Citizens, as a single filter to send those flags.** And in any event, this is a temporary solution until the organisations are onboarded to the relevant channel, and all the same parameters will apply.

Reporting guidelines: For all flags (including this weekend), as we have before, we would appreciate having the following details:

- Name of flagging organisation
- Time of notification (including time of notification via regular tools)
- Platform concerned
- URL to the content
- Language of the content
- Reason for the flag (including *if possible* reference to relevant platform policies)

Finally (and as you know), we recommend that the organisations flag this content via regular in-reporting tools as well.

Many thanks,

Redacted

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Meta

Redacted
Public Policy Manager | EU Affairs
Meta | Mobile Redacted

From: Redacted ec.europa.eu>
Sent: 29 November 2024 11:16
To: Redacted tiktok.com>
Cc: Redacted tiktok.com <Redacted tiktok.com>; Redacted tiktok.com> Redacted Redacted @ec.europa.eu> Redacted ec.europa.eu> Redacted Redacted ec.europa.eu>; Redacted google.com> Redacted <Redacted microsoft.com>; Redacted meta.com>
Subject: RE: [External] RO elections - Rapid Response System

This Message Is From an External Sender

Dear Redacted

Thank you very much for the quick and positive answer on this.

A small point of clarification: as we've called for the activation of the Rapid Response System under the Code we will also invite Elections WG Members to participate to it.

Kind regards,
Redacted

From: Redacted tiktok.com>
Sent: Thursday, November 28, 2024 9:03 PM
To: Redacted @ec.europa.eu>
Cc: Redacted tiktok.com Redacted tiktok.com>; Redacted Redacted ec.europa.eu>; Redacted ec.europa.eu> Redacted (CNECT) Redacted @ec.europa.eu> Redacted google.com> Redacted <Redacted microsoft.com>
Subject: Re: [External] RO elections - Rapid Response System

Hi Redacted

We are happy to activate the RRS mechanism for Romania until December 12th.

In accordance with prior activations, we kindly ask EDMO's BROD hub to use our TikTok RRS channel: Redacted Redacted @tiktok.com

Thank you for facilitating this process.

Kind regards,
Redacted

From: Redacted ec.europa.eu>
Date: Thu, Nov 28, 2024, 18:14

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Subject: [External] RO elections - Rapid Response System
To: Redacted meta.com>, Redacted bytedance.com>,
"Redacted tiktok.com"<Redacted tiktok.com>, Redacted google.com>, Redacted
Redacted google.com>, Redacted google.com>,
"Redacted google.com"<Redacted google.com>, Redacted microsoft.com>, Redacted
Redacted google.com>
Cc: Redacted ec.europa.eu>, Redacted ec.europa.eu>,
Redacted ec.europa.eu>

Dear All,

As you are certainly aware, the first round of presidential elections in Romania has raised concerns regarding the use of social media and the integrity of the electoral process. Following the decision of the Constitutional Court to ask for a recounting of votes, we expect the online space to become even more polarized and the disinformation landscape to potentially shift rapidly.

Building on the positive experience regarding the usefulness of the RSS for the EP and Moldovan elections – and upon requests from local factcheckers, in particular BROD - we are calling for activating the Rapid Response System of the Code.

EDMO's BROD hub covering Romania will be in an excellent position to contribute to the RSS with expertise on the ground, leveraging also its experience and learnings from the Moldovan elections (FunkyCitizens being co-coordinator).

We would propose the RRS to run for 2 weeks (until Wednesday 12 December) to cover the parliamentary elections and the second round of the presidential elections. The first meeting would take place next week.

In view of the urgency of the situation we would kindly ask you to get back to us swiftly on this, but in any event before Friday COB.

We stay at your disposal should you have any question and looking forward to your answer,

Kind regards,
Redacted on behalf of Redacted and the team

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 264

From: [Redacted]tiktok.com]
Sent: 11/29/2024 8:17:20 PM
To: WEZENBEEK Rita [Rita.WEZENBEEK@ec.europa.eu]
CC: [Redacted]roaep.ro; [Redacted]ancom.ro; [Redacted]roaep.ro; [Redacted]roaep.ro; [Redacted]ancom.ro; [Redacted]ancom.ro; [Redacted]tiktok.com; CNECT-DSA-REGISTRY@ec.europa.eu; cnect-dsa-elections@ec.europa.eu; [Redacted]@ec.europa.eu]; [Redacted]@tiktok.com]; [Redacted]@ec.europa.eu]
Subject: Re: [External] RE: TikTok follow up to roundtable on Romanian elections
Attachments: Letter to the European Commission re. Romanian Elections - 29 November 2024.pdf

Dear [Redacted]

Thank you for your confirmation of the additional time - please find attached our response to the questions arising from the meeting on 27 November 2024.

We appreciate that an RFI has now been issued and we are now working on that, but if it would be helpful to have any further discussions, please do let us know.

Best

[Redacted]

Redacted
Email: [Redacted]tiktok.com

On Fri, Nov 29, 2024 at 5:41 PM [Redacted]@ec.europa.eu> wrote:

> Dear [Redacted]
>
> Thank you for your message and thank you for TikTok's participation in today's Round Table.
>
> We look forward to receiving TikTok's reply by end of today,
>
> Best regards,
>
> **Redacted**
>
> **From:** [Redacted]tiktok.com>
> **Sent:** Friday, November 29, 2024 3:04 PM

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

To: Redacted @roaep.ro
Cc: Redacted @ec.europa.eu>; Redacted ancom.ro; Redacted roaep.ro; Redacted roaep.ro; Redacted ancom.ro; Redacted ancom.ro; Redacted @tiktok.com; CNECT DSA REGISTRY <CNECT-DSA-REGISTRY@ec.europa.eu>; CNECT DSA ELECTIONS <cnect-dsa-elections@ec.europa.eu>; Redacted Redacted ec.europa.eu>; Redacted tiktok.com>; Redacted Redacted ec.europa.eu>
Subject: Re: [External] RE: TikTok follow up to roundtable on Romanian elections

Dear Redacted and team

I wanted to update you that we are working to address the questions raised in this thread as best we can in the time available. However, given that a number of key people working on the responses will be attending this afternoon's roundtable, I wanted to let you know that we will need some additional time this afternoon/evening to complete our responses. We shall endeavour to do so as soon as possible after the roundtable finishes.

Yours sincerely

Redacted

Redacted

Email: Redacted tiktok.com

On Thu, Nov 28, 2024 at 9:59 PM Redacted @roaep.ro> wrote:

> Hi,
>
> I want to detail question number 2 to understand the context and TikTok's ineffective response to our request based on a decision by the Central Electoral Bureau.:
>
> 2. Questions for TikTok and the Romanian Authorities - regarding the complaint sent to TikTok on 21 November 2024 and responded to on 22 November 2024 (the '21 November Complaint'):
>
> On 21 November at 08:30 AM, I sent an email to TikTok in order to "remove electoral propaganda materials from the online environment depicting the candidate Calin Georgescu in the 2024 Romanian Presidential elections, which do not contain the financial identification code of the representative" according to the Central Electoral Bureau Decision 175D.
>
> The following response came from TikTok: "

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

• In connection with the Decision of the Central Electoral Bureau ("Biroul Decision") dated 20 November 2024 communicated to us by the Permanent Electoral Authority on 21 November 2024, we confirm that we have removed access in Romania to the video content published by the candidate's account, as specifically referenced in your email below. We do not consider that supporter's accounts are within the scope of the Bureau Decision.

• In connection with the Decision of the Electoral Bureau of County Constituency No. 3 Argeș ("Argeș Decision") dated 21 November 2024 communicated to us by the Permanent Electoral Authority on 21 November 2024, we confirm that we have removed access in Romania to the video content published by the candidate's account, as specifically referenced in your email below. "

What I can confirm is that the first decision was ineffective implemented. I have not seen any materials removed from the respective account and erroneously it is said that access to the video content published by the candidate's account was removed in Romania (the BEC decision is mandatory not only in Romania. It is very difficult to assess what was removed considering that hundreds of accounts, possibly from Romania and abroad, were distributing a post expressly mentioned as a post that violates the legislation and the BEC decision as an example in the email to TikTok). Just yesterday during the ad hoc online meeting, I checked the account and it has dozens of posts not marked as a candidate or with a financial representative code.

Only a small part of the materials were actually removed only on Monday morning, after the end of the vote.

No materials were removed from November 21 until this Monday, but dozens of other materials followed also in the silence period (Saturday and Sunday - voting day). And at this moment there are numerous materials without the financial representative code on this account or distributed massively on other accounts.

Furthermore, on November 22 at 08:37 PM, the platform was notified of decision 39 of the Electoral Bureau for Foreigners in which the bureau ordered the removal of a sound (song) from the entire TikTok platform. We have no response from TikTok to this decision, whether or not they removed the sound, but I can confirm that it is still present on the TikTok network.

As positive aspects, I mention that this evening I received an email from TikTok with a proposal to use the TikTok Safety Enforcement Tool (TSET) – a dedicated online platform that allows to flag content effectively, but I don't know if it will make a difference if Romanian legislation or the Decisions of the Central Electoral Bureau are interpreted arbitrarily.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Best Regards,

Redacted

General Directorate of Informatics

Permanent Electoral Authority

Redacted | Redacted

Redacted roaep.ro | www.roaep.ro

Str. Stavropoleos nr. 6| București - sect. 3

From: Redacted @ec.europa.eu>
Sent: Thursday, November 28, 2024 1:39 PM
To: Redacted ancom.ro; Redacted @roaep.ro; Redacted @roaep.ro
Cc: Redacted @ancom.ro; Redacted ancom.ro; Redacted @roaep.ro; Redacted @tiktok.com; CNECT-DSA-REGISTRY@ec.europa.eu; cnect-dsa-elections@ec.europa.eu; Redacted <Redacted ec.europa.eu>; Redacted tiktok.com>; Redacted Redacted ec.europa.eu>
Subject: FW: TikTok follow up to roundtable on Romanian elections

Dear President Redacted

Dear Vice-President Redacted

Dear Vice-President Redacted

I would like to sincerely thank you and your teams again for participating in yesterday's ad hoc meeting on a first assessment of issues signalled in media and expert reports regarding the potential influence of content published and amplified on TikTok's social media platform on the preliminary outcome of Romania's Presidential Elections.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

As discussed yesterday, I herewith send you the topics and questions identified by TikTok. We have added one question that was not yet listed below in yellow.

We would be very grateful if you could review the issues and inform us whether you see issues that were discussed yesterday but have not been signalled yet, ideally **by today 16.00hrs**. TikTok can then provide answers which, in view of the on-going elections, should come at their earliest convenience; we would propose by **tomorrow 29 November cob at the latest**.

In addition, TikTok has provided a single point of contact for any issues regarding the elections as they proceed.

I would suggest that you perhaps also provide names of contact persons within your authority to whom TikTok representatives can reach out in case of questions or need for clarifications. TikTok may also want to reach out to you to establish higher level contacts in case urgent issues arise.

In addition, I would like to announce that tomorrow, the Commission is organising a Round Table bringing together platforms, the Romanian authorities and civil society organisations to discuss the Romanian elections. An invitation to participate will reach you separately.

Thank you very much for your attention paid to this issue and I propose that we stay in close contact,

Best regards,

Redacted

Redacted

European Commission

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Directorate-General for Communication Networks Content and Technology

'Copernicus' building

L-51 – 4/028

Rue de la Loi 51
B-1000 Brussels/Belgium

Tel. Redacted

Redacted @ec.europa.eu

From: Redacted @tiktok.com>
Sent: Wednesday, November 27, 2024 8:21 PM
To: Redacted @ec.europa.eu>
Cc: CNECT DSA REGISTRY <CNECT-DSA-REGISTRY@ec.europa.eu>; CNECT DSA ELECTIONS <cnect-dsa-elections@ec.europa.eu>; Redacted @ec.europa.eu>; Redacted tiktok.com>
Subject: TikTok follow up to roundtable on Romanian elections

Dear Redacted and team,

Thank you very much for organising the call today to discuss the alleged issues relating to the elections in Romania. As we mentioned on the call, we treat election integrity and these matters with the utmost seriousness, especially in terms of the impact on democracy and role that the integrity of our platform plays in this regard.

We take this opportunity to address two points:

1. How to contact TikTok - to ensure clarity in this regard on the part of the authorities; and
2. Setting out the questions that we understand have been raised to date.

How to contact TikTok

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Taking on board the feedback received today and to ease communications, we would ask that the Romanian authorities use our DSA single point of contact for any issues as the election period proceeds: **DSA-single-point-of-contact-authorities@TikTok.com**

In addition, if the Electoral Authority would like to be onboarded to our TikTok Safety Enforcement Tool (TSET) platform, which will also allow it to flag content, we can arrange that. We look forward to hearing if this would be of interest.

Questions raised

As discussed at the meeting, set out below is a list of the questions that have been put to us by the European Commission and the Romanian authorities. If we have missed anything from the list, please do let us know. In particular, we appreciate that you had a number of questions on the Commission side and, as we did not have these in advance of the meeting, please do let us know if the list below does not record them in full.

We are also in the process of investigating the letter referenced by **Redacted** from the Romanian Audiovisual Authority, which was stated to have been sent to TikTok. This was not a document that those of us on the call were aware of, and while we investigate if it has been received, it would be greatly appreciated if you could ask them (i) to confirm when and to whom it was sent and (ii) to provide a copy in the meantime in order to avoid any undue delays.

Questions / action items raised on the call

1. **Question for TikTok** - regarding the Romanian delegation that visited TikTok's Transparency & Accountability Centre in October 2024, TikTok is to set out who attended and what topics were covered.
2. **Questions for TikTok and the Romanian Authorities** - regarding the complaint sent to TikTok on 21 November 2024 and responded to on 22 November 2024 (the **'21 November Complaint'**):

    1. both TikTok and the Romanian authorities are to set out their approach to this complaint and the legal basis for their positions (plus clarify response times for this particular complaint); and
    2. in respect of the content that TikTok stated was geo-blocked in response to the 21 November Complaint, the Romanian authorities will provide the URLs to the content that they have said remains available in Romania notwithstanding TikTok's position that the content posted by politicians / political parties was geo-blocked.

3. **Question for the Romanian Authorities** - regarding responses more generally, the Romanian Authorities are to provide any communications that were sent and which they

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

consider were not responded to in a timely manner, so TikTok can look into these to see what happened and establish if any process changes are required.

4. **Questions from the Romanian authorities** - TikTok to explain how it engages with experts and local Member State authorities, including under the DSA, to address the assertion made that it is not paying due attention to relevant local authorities (e.g Ministry of Justice) who are actually responsible for administering the election and have local knowledge and understanding local legal requirements (e.g during 'silence periods').
5. **Question from the Commission** - seek an explanation from TikTok as to how to assess a situation where one candidate has had exponential growth in views.

    1. Related question from the Romanian Audiovisual Authority asking for responses from TikTok to a letter and enclosed document which sets out the questions it has regarding the increase in growth for the candidate in question. [**TikTok Note:** *as flagged above, TikTok is looking for this letter but kindly asks that the Commission provide it in the meantime if they have it or are able to obtain a copy.*]
    2. Related Question from Romanian authorities: Please explain how TT's algorithm worked in the past two months in Romania in terms of amplifying content provided by the candidates or related to their discourse.

6. **Question from the Commission:** asking TikTok to set out its standard policy regarding the prohibition on political ads and whether the policy was effectively enforced in the Romanian election - in particular, have we identified such ads in Romania? [**TikTok Note:** as mentioned on the call, we do collect this data and it is typically produced, validated and shared after the relevant election].

7. **Question from the Commission:** regarding the potential role of paid influencers in the growth of this candidate, can TikTok set out its assessment of any such role (e.g to establish If they have been paid to promote content/the candidate)?

Thank you again, and kind regards,

Redacted

--

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted

Redacted
Email: Redacted tiktok.com

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 265

TikTok Technology Limited
Registered Address
The Sorting Office,
Ropemaker Place,
Dublin 2,
D02 HD23
Email: Redacted tiktok.com

From: **Redacted**
To [CNECT-DSA-Registry@ec.europa.eu]

CONFIDENTIAL

29 November 2024

Dear **Redacted**

We write further to the meeting that took place between the European Commission (the "**Commission**"), the National Authority for Management and Regulation in Communications (the Romanian Digital Services Coordinator), the Permanent Electoral Authority of Romania and the National Audiovisual Council (together, the "**Romanian Authorities**") on 27 November 2024 (the "**27 November 2024 Meeting**"). We also refer to the subsequent email correspondence between us, setting out an agreed list of questions arising from the 27 November 2024 Meeting for TikTok to answer by today. The answers to these questions are at **Appendix 1** to this letter.

We welcome the engagement from the Commission and the Romanian Authorities on this important matter and we are working hard to address the concerns raised to date. To that end, we have sought to provide as much information as we can in the limited period between receiving the list of questions and today's deadline.

As you will appreciate, this is a highly dynamic situation – and this fact, together with the time constraints, mean that in some instances we have not been able to address every point raised within every question. We are also conscious that the Commission has now issued a formal Request for Information ("'**RFI**") under Article 67 DSA in respect of this matter. As such, we trust that the Commission and the Romanian Authorities will accept that the responses provided in Appendix 1 constitute a good faith effort on TikTok's part to respond to the immediate points that have been raised to the best of its ability, and that further and/or updated responses, where relevant, will also now be provided in response to the RFI.

We are, of course, happy to continue a dialogue (in parallel to responding to the RFI) on any topics that need to be addressed more immediately. In this regard, we propose that any material issues are raised in our existing email thread, and that more operational points are raised via TikTok's designated DSA single point of contact email address (i.e. DSA-single-point-of-contact-authorities@TikTok.com).

If there are any points that you or the Romanian Authorities would like to discuss further in light of TikTok's responses in Appendix 1 (to the extent they cannot be addressed via the RFI response), please do not hesitate to let us know.

Yours sincerely

Redacted

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TikTok

Appendix 1

Question 1

Regarding the Romanian delegation that visited TikTok's Transparency & Accountability Centre in October 2024, TikTok is to set out who attended and what topics were covered.

This meeting took place on 2 October 2024. The Romanian delegation comprised the following attendees:

- Redacted – Minister of Research, Innovation and Digitalisation;
- Redacted – Counselor, Ministry of Research, Innovation and Digitalisation (the "**MCID**");
- Redacted – Director, MCID; and
- Redacted – Minister Counselor, Ministry of Foreign Affairs (Embassy of Romania in Dublin).

During the visit, the attendees were given a tour of TikTok's 'Transparency and Accountability Centre' (the "**TAC**"), which recently opened in Dublin.[1] The TAC allows regulators, policymakers, politicians, researchers, academics, businesses and many other expert audiences from around the world to see first-hand how teams at TikTok undertake the critically important work of securing our community's safety, data and privacy.

In addition to the TAC tour, the following topics were discussed with the Romanian delegation:

- how TikTok ensures the integrity of elections and media literacy campaigns to educate users; and
- countering misinformation (which TikTok defines as false or misleading content, whether it is intentional "disinformation" or otherwise) – in particular Russian misinformation.

Question 2

*Regarding the complaint sent to TikTok on 21 November 2024 and responded to on 22 November 2024 (the '**21 November Complaint**'), TikTok is to:*

a. *set out its approach to this complaint and the legal basis for their positions (plus clarify response times for this particular complaint); and*

b. *in respect of the content that TikTok stated was geo-blocked in response to the 21 November Complaint, the Romanian authorities will provide the URLs to the content that they have said remains available in Romania notwithstanding TikTok's position that the content posted by politicians / political parties was geo-blocked.*

In response to part (a) above:

Relevant timelines regarding the 21 November Complaint

- **First Email re: Decision 175D:** TikTok received an email signed off by the General Manager, General Directorate of Informatics at Romania's Permanent Electoral Authority at 7:30am (GMT) / 9:30am (Romania time) on Thursday 21 November 2024 (the "**First Email re: Decision 175D**"). The email attached the decision of the "Biroul Electoral Central" ("**Decision 175D**") and listed two TikTok accounts.

[1] See https://newsroom.tiktok.com/en-eu/dubtac

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

- **Second Email re: Decision 175D:** TikTok received a further email signed off by the General Manager, General Directorate of Informatics at Romania's Permanent Electoral Authority at 7:58pm (GMT) / 9:58pm (Romania time) on 21 November 2024 (the "**Second Email re: Decision 175D**"), which formed part of the same email thread as the First Email re: Decision 175D (and which again attached a copy of Decision 175D). The Second Email re: Decision 175D listed: (i) 14 video URLs relating to the "candidate account"; and (ii) a further 6 video URLs relating to "supporter accounts".
    - TikTok reviewed the 14 videos in accordance with its internal frameworks for assessing content removal requests, and undertook a legal assessment involving external legal advisors.
    - Based on the external assessment of Romanian law, TikTok concluded that the 14 videos reported in the Second Email re: Decision 175D were likely in violation of Decision 175D, and therefore unlawful under Romanian law.
    - On this basis, access to those 14 videos was blocked in Romania by 9:51am (GMT) / 11:51am (Romania time) on Friday 22 November 2024.
    - As regards the 6 videos relating to "supporter accounts" listed in the Second Email re: Decision 175D, based on the external assessment of Romanian law TikTok concluded that these 6 videos were outside the scope of Decision 175D, and therefore not unlawful under Romanian law.
- **Follow-up Email re: Decision 175D:** TikTok received a further email (again signed off by the General Manager, General Directorate of Informatics at Romania's Permanent Electoral Authority) at 10:16am (GMT) / 12:16pm (Romania time) on Thursday 28 November 2024 (the "**Follow-up Email re: Decision 175D**"). It stated: "*Please also tell us how you have implemented Decision 175D of the Central Electoral Bureau which expressly states that all materials containing Calin Georgescu images must be removed*". TikTok responded to the Follow-up Email re: Decision 175D today, providing a summary of the actions taken in response to Decision 175D.

Legal basis for TikTok's approach to handling of the 21 November Complaint

- TikTok is deeply committed to election integrity, and has policies and processes in place to handle government content removal requests in compliance with applicable laws and consistent with TikTok's commitments to human rights principles. This necessarily requires TikTok's teams to carefully assess any government content removal requests on a case-by-case basis, which is especially important in the context of elections.
- Accordingly, TikTok conducted a legal assessment of the requests sent by the Permanent Electoral Authority and, based on that legal assessment, took action to block access in Romania to the 14 videos that TikTok concluded were likely in violation of Decision 175D, and therefore unlawful under Romanian law.
- In this regard, TikTok considers that it acted in accordance with both the letter and spirit of European law in its handling of the 21 November Complaint:
    - Article 8 of the DSA prohibits any national authority from imposing a general obligation to monitor or to actively seek facts or circumstances indicating illegal activity. Therefore, neither the Biroul Electoral Central nor the Permanent Electoral Authority can impose such an obligation on TikTok. Instead, it is appropriate for individuals or entities to submit specific reports of allegedly illegal content to TikTok and/or for authorities to issue orders to act against specific items of illegal content.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

- In this regard, Article 9 of the DSA also sets down conditions that authorities must meet when issuing orders to act against specific items of illegal content, to ensure that such orders are precise, targeted and well-substantiated. The 21 November Complaint did not comply with these conditions.

- Based on the external assessment of Romanian law, and in light of the Permanent Electoral Authority's request, we note the following regarding TikTok's interpretation of Decision 175D:
  - **Accounts / profiles:** The existence of a TikTok account, in itself, should not be considered "propaganda material" within the scope of Decision 175D (which is why the two accounts listed in the First Email re: Decision 175D were not restricted or removed). Nonetheless, one of the accounts was subsequently banned for impersonation in violation of TikTok's Community Guidelines.
  - **Request to remove all images of Calin Georgescu:** The statement in the Follow-up Email re: Decision 175D, to the effect that Decision 175D "*expressly states that all materials containing Calin Georgescu images must be removed*", does not appear to reflect the scope of Decision 175D.
- Further, we note that Decision 175D is not addressed to TikTok and does not reference TikTok in any manner, but is instead a decision of general application. Moreover, Decision 175D does not identify any specific content to be removed.

In response to part (b) above (i.e. "*in respect of the content that TikTok stated was geo-blocked in response to the 21 November Complaint, the Romanian authorities will provide the URLs to the content that they have said remains available in Romania notwithstanding TikTok's position that the content posted by politicians / political parties was geo-blocked*"), TikTok notes that these URLs have not been provided in any subsequent emails on the existing thread with the Commission and the Romanian Authorities. If the URLs are provided, TikTok would be happy to review them in order to establish whether or not the intended geo-block was applied.

Question 3

TikTok to explain how it engages with experts and local Member State authorities, including under the DSA, to address the assertion made that it is not paying due attention to relevant local authorities (e.g. Ministry of Justice) who are actually responsible for administering the election and have local knowledge and understanding of local legal requirements (e.g. during 'silence periods').

TikTok engages with local Member State authorities and elections experts in relation to elections in multiple ways. This is done primarily through engagement via TikTok's Government Relations (GR) teams and Outreach Partnership Management (OPM) teams, as well as via TikTok's Trust and Safety, Legal and DSA Compliance functions.

To ensure that TikTok's Trust and Safety teams are as fully prepared as possible for upcoming elections, TikTok invites its external fact-checking partners to attend a 'Speaker Series' session with TikTok's internal teams. During this session, fact-checking experts present key potential risks and challenges during the election period directly to the TikTok'sTrust and Safety teams responsible for monitoring the election.

These sessions play a key role in TikTok's internal election-related preparations, by ensuring that all relevant internal teams are abreast of ongoing risks during the election period. For the Romanian elections specifically, the session was conducted with TikTok's Media

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Literacy Partner in Romania, Funky Citizens.[2] This is a continuation of TikTok's collaboration with Funky Citizens, which previously provided a Speaker Series for TikTok in May 2024 with insights relating to the EU elections. TikTok has also collaborated with Funky Citizens to produce media literacy content in Romanian, giving users tips to combat misinformation and disinformation.

In addition, with respect to the Romanian elections in particular, TikTok organised or participated in the following sessions and calls:

1. 13 May 2024 – Call regarding elections integrity and deepfakes reporting system

   On 13 May 2024, TikTok attended a call with the MCID (which was also attended by Meta and Google) to discuss reporting of deepfakes.

2. 3 June 2024 – Call regarding elections integrity and deepfakes reporting system

   On 3 June 2024, TikTok attended a call with the MCID. During the call, attendees discussed local and European elections integrity, reporting of deepfakes, and access to our 'TikTok Safety Enforcement Tool' (TSET), which facilitates the reporting of content.

3. 4 June 2024 – Elections integrity and Project Clover meeting

   On 4 June 2024, TikTok attended a meeting in Bucharest with the MCID to discuss elections integrity, GPPPAs[3] and Project Clover.

4. 2 October 2024 – Visit to the TAC by the Minister of Research, Innovation and Digitalisation and others

   Please see TikTok's response to Question 1 above.

5. 11 October 2024 – Further meeting with Funky Citizens

   A final online meeting with Funky Citizens was held on 11 October 2024, to agree on a joint media literacy campaign, which was initiated on 22 November 2024.

6. 22 October 2024 – Elections integrity call with the Romanian Digital Services Coordinator

   TikTok's Government Relations team proactively organised election-dedicated meetings with ANCOM, Romania's Permanent Electoral Authority and the MCID to establish points of contact before the elections, to offer access to TikTok's reporting tools and propose joint cooperation in creating an Election Centre for the Romanian election. Election Centres are special sections on the TikTok platform that TikTok has developed and makes available to users ahead of key elections – working with local fact-checkers, electoral commissions and others – to connect users with reliable elections and media literacy information, including voting-related information.

7. Between 29 October and 7 November 2024 – Online sessions for GPPPAs

   TikTok's Government Relations team organised dedicated sessions with political groups in Romania, to inform them about TikTok's elections-related policies. This

---

[2] https://funky.ong/en/despre-noi/

[3] TikTok classifies certain political accounts as "*Government, Politician and Political Party Accounts*" ("**GPPPAs**"). This is explained further here: https://support.tiktok.com/en/using-tiktok/growing-your-audience/government-politician-and-political-party-accounts

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013837

included educating key political entities and persons about safety measures. These sessions took place on:

- 29 October 2024, with the Social Democratic Party (PSD);
- 4 November 2024, with the Alliance for the Union of Romanians (AUR);
- 5 November 2024, with the Renewing Romania's European Project Party (REPER);
- 5 November 2024, with the Save Romania Union Party (USR);
- 7 November 2024, with the Democratic Alliance of Hungarians in Romania (UDMR); and
- 7 November 2024, with the United Right Alliance (FD and PMP).

8. 14 November 2024 – Influence Operations Intelligence and Discovery session in Bucharest and media briefing

   On November 14, TikTok in cooperation with the MCID, organised an onsite session on disrupting covert influence operations ("**CIOs**"). The meeting was attended by local Romanian experts from the MCID, ANCOM, DNSC and Police. Representatives from TikTok's Influence Operations Intelligence and Discovery and Global Influence Operations Threat Disruption teams explained TikTok's approach to these risks and shared their professional knowledge.

Question 4

Explain how TikTok assesses a situation where one candidate has had exponential growth in views?

a. *Related question from the Romanian Audiovisual Authority asking for responses from TikTok to a letter and enclosed document which sets out the questions it has regarding the increase in growth for the candidate in question.*

TikTok has been unable to locate a copy of the letter (and enclosure) from the Romanian Audiovisual Authority referred to in this question. A copy was also requested when TikTok provided the list of questions on 27 November 2024 (see TikTok's email to the Commission and the Romanian Authorities at 20:21 CET) but was not provided in any of the subsequent replies.

We would be happy to respond to this question if these documents can be provided to us – e.g., by sending them to TikTok's designated DSA single point of contact email address[4] or via the aforementioned email thread.

b. *Related Question from Romanian authorities: Please explain how TT's algorithm worked in the past two months in Romania in terms of amplifying content provided by the candidates or related to their discourse.*

TikTok's recommender systems do not specifically amplify content from GPPPAs (or other political accounts), as distinct from other accounts or categories of content. They are designed to offer TikTok's users a personalised experience, by predicting and recommending to each user the content that the user might find relevant and interesting.

The three main factors used by TikTok's recommender systems are: (i) user interactions (e.g., what content users like, share, comment on, and watch); (ii) content information (e.g., sounds, hashtags, number of views, and the country from which the content was published); and (iii) user information (e.g., device settings, language preference, and non-precise location details). Recommended content can be viewed by users through several feeds and

[4] i.e., DSA-single-point-of-contact-authorities@TikTok.com

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

features on the platform, including: For You Feed ("**FYF**"); Following feed; LIVE; Search; and Comments. The recommender systems produce different outputs for each feed/feature.[5]

All videos uploaded to TikTok are assessed by TikTok's automated and/or human content moderation processes. Content that is assessed as violating TikTok's Community Guidelines (including the "*Misinformation*" and "*Civic and Election Integrity*" policies under the "*Integrity and Authenticity*" section of the Community Guidelines[6]) are removed and therefore not available for recommendation. The reach of other content, such as elections-related claims which may amount to Election Misinformation but cannot be verified as such, may be limited – e.g., by being made ineligible for recommendation on users' FYF, the main feed through which users discover content.

To complement and supplement TikTok's content moderation processes, if a video rises in popularity (i.e. if it has an increasing number of views), it is flagged at stages for further automated, and potentially manual, review. This mitigates the risk of Election Misinformation being recommended in a user's FYF (or going 'viral'). As part of this process, TikTok's automated systems take into account various factors – including video views, number of likes, comments and user reports on the video, etc – using specialised models to assess the video's degree of risk (i.e. the likelihood that the video contains content that violates the Community Guidelines) and predicted video view count growth.

Question 5

Set out your standard policy re. the prohibition on political ads and whether the policy was effectively enforced in the Romanian election - in particular, have you identified such ads in Romania?

TikTok's standard policy concerning political ads is set out in Schedule 1.

TikTok is gathering more complete data relevant to the effectiveness of its policy enforcement in the context of the Romanian election to enable it to respond to this question with confidence, and will address this question further in its RFI response.

In the meantime, TikTok is able to offer the following information and metrics indicating the effectiveness of its ads policy enforcement in the EU.

Political ads represent a relatively small proportion of the violative ads content that TikTok actions on the platform. The table below sets out for the EU the number of ads which were uploaded in the period January to end of June 2024, the number and percentage which were rejected or removed for violating TikTok's policies and the number and percentage of those rejections and removals which relate to political ads.

Total ads uploaded in EU (Jan to end June 2024)	Total ads rejected/ removed for any policy violation	% Rejected/ removed for any policy violation	Total rejections/ removals which relate to political ads	% of rejections/ removals which relate to political ads
4,330,715	315,937	7.3%	1,232	0.39%

[5] For example, users' Following feed shows content from accounts that the user has chosen to follow.

[6] https://www.tiktok.com/community-guidelines/en/integrity-authenticity

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013839

In addition, the overwhelming majority of political ads submitted in violation of TikTok's political advertising policies are rejected at the pre-publication stage and therefore do not appear on the platform. Indeed, in Q2 2024, 98.36% of the political ads intended for the EU market were rejected or removed before receiving any views (i.e. under 2% of the ads that were ultimately identified as political ads received any views).

Question 6

Regarding the potential role of paid influencers in the growth of this candidate, can TikTok set out its assessment of any such role (e.g to est. If they've been paid to promote content/the candidate)?

As explained in response to Question 4 above, the "*Civic and Election Integrity*" section of TikTok's Community Guidelines outlines that TikTok does not allow paid political promotions.

TikTok's political advertising policy covers both traditional paid advertisements and creators who receive compensation to support or oppose a candidate for office within organic content.

In addition, TikTok prohibits political Branded Content (i.e. content promoted in exchange for payment or any other incentive). Further details of this policy are included in Schedule 1.

With respect to the growth of this Romanian candidate in particular, on Friday, 29 November 2024 (today) we removed a network of 78 accounts for violating our policy against covert influence operations. We assess that this network operated from Romania and targeted a Romanian audience. The individuals behind this network created inauthentic accounts in order to promote candidate Călin Georgescu on TikTok.

(This is the only covert influence network disruption that we have made in Romania to date, where we have found Călin Georgescu to be a prominent topic and we currently do not assess this particular network to be associated with paid influencers. Details surrounding this network will be disclosed in our November 2024 Covert Influence Operations transparency report).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Schedule 1

TikTok's approach to political ads

On TikTok, users can share political content organically as long as it is in line with our Community Guidelines. However, TikTok does not allow political content to feature in ads. As detailed below, TikTok implements and communicates this prohibition on political ads publicly through relevant user terms and conditions in clear, plain, intelligible, user-friendly and unambiguous language. TikTok also maintains a range of proactive and reactive measures and processes to detect and remove political ads from the TikTok platform when attempts are made to upload them. Although TikTok seeks to continuously improve its content related systems and processes, TikTok is confident that it has reasonable and proportionate moderation measures in place with respect to content which violates these political advertising policies.

TikTok does not allow ads featuring political content across any of its monetisation features, including paid ads, creators being paid to make branded political content, and other promotional tools on the platform. TikTok also imposes prohibitions at the account level for advertisers it labels as GPPPAs[7], who are not able to create advertising accounts or access advertising features.

TikTok's policies and relevant Terms and Conditions, on political advertising, and what type of political ads are not allowed on the platform.

(a) Policies and Relevant Terms and Conditions

TikTok does not allow political advertising on the TikTok platform. This prohibition applies both to "advertisements", as defined under the DSA, where TikTok receives remuneration itself, and content for which TikTok is not remunerated but the uploader is themselves paid, or receives a benefit-in-kind, for posting the relevant content. For the purpose of this response, we have focused on advertising of a political nature where TikTok itself receives remuneration.

TikTok sets out in detail below its standard policies and relevant terms and conditions on Political Ads, including relevant quoted extracts in italics.

TikTok For Business Commercial Terms of Service

- *2.1 Certain Commercial Products or features of Commercial Products are governed by additional terms, guidelines, rules or policies ("**Supplemental Terms**"). Supplemental Terms include the linked terms set out below, along with any Supplemental Terms that TikTok incorporates by reference elsewhere in these Commercial Terms or makes available to you when using a particular Commercial Product or feature of a Commercial Product.*
  - *(a) If you use Commercial Products to access or purchase advertising inventory or advertising-related services, you agree to the Advertising Terms and Advertising Policies.*
  - *(b) …*
  - *(c) …*

---

[7] https://support.tiktok.com/en/using-tiktok/growing-your-audience/government-politician-and-political-party-accounts

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

(d) *If you use Commercial Products to share content or interact with TikTok or others, you agree to the Community Guidelines and the Branded Content Policy.*

TikTok For Business Advertising Terms

- *1.1 These TikTok For Business Advertising Terms (these "**Advertising Terms**") govern your use of the Commercial Products (as defined in the Commercial Terms) to submit, create, or place advertising or any other type of sponsored or promotional content , or to plan or manage ad campaigns, which might include the use of creators or other third parties ("**Ad Services**").*

- *1.3 These Advertising Terms form part of the Commercial Terms. They apply in addition to the Commercial Terms and any other applicable Supplemental Terms (as defined in the Commercial Terms). To the extent any specific provision of these Advertising Terms conflicts or is inconsistent with any term of the Commercial Terms, the specific provision of these Advertising Terms will prevail.*

- *3.1 You will ensure any of Your Content (as defined in the Commercial Terms) that is provided for Ad Services ("**Ad Content**") will comply with the Advertising Policies.*

Advertising Policies **("Advertising Policies")**

- Advertising on TikTok - First Things to Note

  1. Overview

  As an advertiser on TikTok, you are responsible for your ads. This means complying with all applicable laws, rules and regulations that apply to the targeted regions (including any advertising codes, such as the CAP Non-broadcast Code in the UK), creating and promoting honest, safe, and respectful ads. Additionally, you must comply with our Terms of Service, Community Guidelines, and all of TikTok's policies found in our Business Help Center governing the use of our services. We strive to keep up with changing times and are committed to providing a better user experience when viewing TikTok Ads.

- Politics, Religion and Culture

  Principle

  At TikTok, users can share political content organically as long as it's in line with our Community Guidelines. However, we do not allow political content to feature in advertising. We also do not allow content that disrespects religious elements, cultures, or figures, or any hateful behavior, hate speech, or the promotion of hateful ideologies.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Political Content

Policy

We do not allow ads featuring political content across any of our monetization features, including paid ads, creators being paid to make branded political content, and other promotional tools on the platform. We also impose prohibitions at the account level for advertisers we identify as politicians and political parties, who are not able to create advertising accounts or access advertising features.

The types of advertisers that are not allowed to advertise on our platform:

- *Candidates or nominees for public office, political parties, elected or appointed government officials*
- *The spouses of candidates, elected, or appointed government officials with official duties or offices*
- *Royal Family members with official government capacities*

Government entities may be eligible to advertise if managed by a TikTok Sales Representative.

Examples of what is not allowed:

- *Referencing, promoting, or opposing candidates or nominees for public office, political parties, or elected or appointed government officials*
- *Advocacy for or against past, current, or proposed referenda, ballot measures, and legislative, judicial, or regulatory outcomes or processes. This includes ads that promote or attack government policies or track records.*
- *Referencing, promoting, or selling merchandise that features prohibited individuals, entities, or content, including campaign slogans, symbols, or logos*
- *References to an election, including voter registration, voter turnout, and appeals for votes, such as ads that encourage people to vote*

Terms of Service

- **4.5 What you can't do on the Platform**

  Our Community Guidelines apply to everyone and to all content on the Platform. If you want to use the Platform, you can't create, post, share, link to or otherwise interact with content in breach of our Community Guidelines.

- **6.2 TikTok's rights**

  In the event of any suspected breach of these Terms or our Community Guidelines, we may investigate. While we do so, we are allowed to take down some or all of your content, or suspend your access to some or all the features of the Platform, acting reasonably and objectively depending on the seriousness of the suspected breach.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Community Guidelines

- **Remove content that we do not allow**

  Everyone who joins TikTok has the ability to freely share content on the platform. However, we remove content--whether posted publicly or privately--when we find that it violates our rules.

- **Commercial Disclosure and Paid Marketing**

  ...

  We do not allow any form of paid political marketing on TikTok. *Learn more about our policy regarding Government, Politician and Political Party accounts.*

- ***Civic and Election Integrity***

  Elections are important events and are often the subject of intense discussion and analysis. We try to balance enabling these discussions, while also being a place that brings people together and does not cause division. ***We do not allow paid political promotion, political advertising, or fundraising by politicians and political parties (for themselves or others).*** *Our political advertising policy includes both traditional paid advertisements and creators receiving compensation to support or oppose a candidate for office.*

  We want to enable the informed exchange of civic ideas in a way that fosters productive dialogue. ***We do not allow misinformation or content about civic and electoral processes that may result in voter interference, disrupt the peaceful transfer of power, or lead to off-platform violence.***

  Content may be ineligible for the FYF if it contains misinformation that can hinder the ability of a voter to make an informed decision. To be cautious, unverified claims about an election and content temporarily under review by fact-checkers may also be ineligible for the FYF.

  To help you manage your TikTok experience, we may apply warning labels to content that has been assessed by our fact-checking partners and cannot be verified as accurate. Learn more about our election integrity work, and Government, Politician and Political Party accounts.

Government, Politician, and Political Party Accounts (GPPPA) Policy

- ***Advertising***

  TikTok has long prohibited political advertising, including both paid ads and creators being paid to make branded political content. This also includes the use of promotional tools available on the platform, like Promote or TikTok Shop. In addition to our political advertising content policy, we also impose prohibitions at the account level. This means

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

that accounts we identify as belonging to politicians and political parties have their access to advertising features turned off.

We recognize that there will be occasions when official government organizations may have a reason to advertise on TikTok, for example, to promote public health initiatives like COVID-19 booster programs. We'll continue to allow government organizations to advertise in limited circumstances, where they're required to work with a TikTok representative.

- ***Campaign (election) fundraising***
  Just as political advertising doesn't fit with our goal of making TikTok a place that brings people together, neither does campaign fundraising. Solicitations for campaign fundraising by GPPPAs are not permitted on the platform. That includes content like a video from a politician asking for donations, or a political party directing people to a donation page on their website.

Branded Content Policy

- ***Political content*** *- All political or issue-based advertising is prohibited:*
  - *Candidates or nominees for public office, political parties, elected or appointed government officials are prohibited from advertising.*
  - *The spouses of candidates, elected, or appointed government officials with official duties or offices are prohibited from advertising.*
  - *Royal Family members with official government capacities are also prohibited from advertising.*
  - *Political and issue-based advertising includes:*
    - *ads that reference, promote, or oppose candidates or nominees for public office, political parties, or elected or appointed government officials;*
    - *ads that reference an election, including voter registration, voter turnout, and appeals for votes;*
    - *ads that include advocacy for or against past, current, or proposed referenda, ballot measures, and legislative, judicial or regulatory outcomes or processes, including ads that promote or attack government policies or track records; and*
    - *ads that reference, promote, sell, merchandise that feature prohibited individuals, entities or content, including campaign slogans, symbols, or logos.*

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 266

TikTok Technology Limited's response to Section 1.1 of RFI
(Ref. Ares(2024)8716531 - 06/12/2024)
7 December 2024

1. TikTok Technology Limited submits this response to ***Section 1.1*** of the Commission's Request for Information Ref.Ares(2024)8716531 (the "**RFI**") as the legal entity that is the provider in the EU of the designated very large online platform TikTok (the "**Platform**") under Regulation (EU) 2022/2065 (the "**DSA**").

Preliminary Points

2. TikTok notes the following preliminary points:

   a. TikTok takes the issues raised in the RFI extremely seriously and, as always, is willing to engage in a constructive dialogue with the Commission on these issues. These matters are under continued investigation and we will follow up with relevant information to the extent we identify it, including in our responses on 13 December 2024.

   b. Following receipt of a courtesy copy received informally after business hours on 5 December 2024, TikTok formally received the RFI through the DSA Registry at 9.53 a.m. CET on 6 December 2024. The Commission requested a response to ***Section 1.1*** of the RFI by the end of the day on 6 December 2024. TikTok appreciates the Commission's subsequent agreement to extend that deadline to 10.00 a.m. CET on 7 December 2024.

   c. TikTok has made best efforts, in the limited time available, to respond to the Commission's questions in ***Section 1.1*** of the RFI. TikTok remains available to discuss its responses to any of these questions.

   d. As TikTok explained to the Commission on a call during the afternoon of 5 December 2024, TikTok does not have human/professional translations of the declassified documents referred to in ***footnote 6*** of the RFI (the "**Declassified Documents**") to which ***Section 1.1*** relates. TikTok appreciates the Commission providing on a courtesy basis, at 3:33 p.m. CET on 6 December 2024, what TikTok understands to be machine translations of the Declassified Documents. However, TikTok notes that such machine translations may lack nuance and clarity as compared to human/professional translations. Indeed, the Declassified Documents are difficult to understand in places, and at times it is unclear what evidence they are referring to.

   e. TikTok is unclear as to the precise scope of the "*findings*" to which ***question 1*** of ***Section 1.1*** relates. ***Question 1*** states that "*these documents*" (i.e., the multiple documents referred to in ***footnote 6*** of the RFI) "*contain a number of findings regarding the use made of TikTok in the context of the Presidential elections*" (the "**Findings**"), and asks TikTok to provide its "*response on each of the assertions made in the intelligence documents*" (also plural). Based on its review of the machine translations provided, TikTok understands that the Findings to which the Commission refers appear in one document – i.e., the Declassified Document at ***footnote 7*** of the RFI ("**Document SRI I**"). To the extent that the Commission would like TikTok to address any other findings, TikTok would appreciate the Commission identifying them expressly.

   f. Further, TikTok observes that, while the Findings have been published, evidence underpinning these Findings has not. Without prejudice to the positions set out herein, TikTok requests the Commission to provide as a matter of urgency all such evidence. This will allow TikTok to properly investigate these matters and come to a more fulsome understanding of, and response to, these allegations.

   g. As the Commission is aware, this is a fast-moving, fluid situation, and TikTok is currently investigating the factual position in relation to the Findings, and the situation more broadly. With these limitations in mind, and subject to them, TikTok provides herein the relevant responsive information that it has been able to obtain in the time available, and in the context of such ongoing investigations.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

h. For completeness, TikTok also notes that the relevant internal teams that have needed to gather and provide the information requested by the Commission in ***Section 1.1*** of the RFI significantly overlap with teams currently assessing and managing risks related to the same, as well as with teams that are simultaneously engaged with preparing responses to RFI Ref. Ares(2024)6974764, RFI Ref. Ares(2024)8523013, and ***Section 1.2*** of this RFI – all of which are due on 13 December 2024.

i. In light of the foregoing circumstances, TikTok respectfully reserves its position regarding the legal and procedural basis for ***Section 1.1*** of the RFI.

j. Further, please note that TikTok has been fully engaged with the Romanian authorities on the issue of Covert Influence Operations ("**CIOs**"). TikTok will provide a list of the meetings it has had with Romanian authorities with its response to ***Section 1.2*** of this RFI.

k. In addition, as the Commission will be aware:

i. Article 8 and Recital 30 DSA expressly state that the DSA does not impose, and nothing in it should be construed as imposing, a general monitoring obligation or a general active fact-finding obligation on providers, nor a general obligation for providers to take proactive measures in relation to illegal content; and

ii. Article 14(4) DSA requires providers to enforce restrictions in their terms and conditions in a "*diligent, objective and proportionate manner*" with due regard to, among other things, users' fundamental rights such as freedom of expression.

TikTok understands that nothing in this RFI should be construed as suggesting otherwise.

Finally, TikTok directs the Commission's attention to a newsroom post published by TikTok at 7:30 p.m. CET on 6 December 2024 (the "**Newsroom Post**"),[1] which provides further information relating to certain assertions made in the Declassified Documents.

Section 1.1 - Questions and Answers

(1) The Romanian President declassified the information submitted by the Romanian Intelligence Service, the Foreign Intelligence Service and the Ministry of Internal Affairs at the meeting of the Supreme Council for the Defence of the Country on 28 November 2024 ([6]). These documents contain a number of findings regarding the use made of TikTok in the context of the Presidential elections held on 24 November 2024, including the following:

- **The existence of inauthentic coordination on the platform, including through the use of hashtags, to promote one presidential candidate;**

3. TikTok understands that this statement ("*The existence of inauthentic coordination on the platform, including through the use of hashtags, to promote one presidential candidate*") refers to the assertion at paragraph 1.1 of Document SRI I that a: "*network of accounts directly associated directly with Călin GEORGESCU's campaign initially consisted of 25,000 accounts on the TikTok platform, which became very active two weeks before the election date*".

4. TikTok has not found, nor been presented with, any evidence of a coordinated network of 25,000 accounts associated with Mr Georgescu's campaign. Per TikTok's standard process, TikTok continues to monitor, investigate, and take action on attempts to deceive its community, and TikTok investigates leads that are reported to it.

[1] See https://newsroom.tiktok.com/en-eu/continuing-to-protect-the-integrity-of-tiktok-during-romanian-elections.

5. TikTok's Integrity & Authenticity Community Guidelines[2] prohibit CIOs.[3] In accordance with TikTok's standard processes for enforcing its Integrity & Authenticity Community Guidelines and countering CIOs, TikTok continues to proactively monitor for, investigate, and take action against identified attempts to deceptively manipulate the Platform. In the context of the Romanian elections, TikTok has identified and disrupted the following CIOs (as explained in the Newsroom Post[4]):

   a. **A network of 22 accounts operating from Romania** that attempted to target Romanian audiences in order to spread misinformation and promote narratives critical of the current Romanian government, which TikTok transparently reported in September 2024.[5]

   b. **A network of 78 accounts operating from Romania**, with 1,781 followers, that attempted to promote Mr Georgescu on TikTok. To date, this is the only CIO network that TikTok has identified that focused on promoting Mr Georgescu as a prominent topic. TikTok assessed that this CIO network was operating from Romania and disrupted it in late November 2024. This will be reflected in TikTok's November Covert Influence Operations Report.

   c. **A network of 12 accounts operating from Romania**, with 1,704 followers, that attempted to amplify narratives in Romania favorable to independent political candidate Mircea Geoană. TikTok assessed that this CIO network was operating from Romania and disrupted it in late November 2024. This will also be reflected in TikTok's November Covert Influence Operations Report.

   d. In addition to CIO networks specifically originating from Romania, TikTok notes that it also operates broader defences against covert influence campaigns across Europe. In September 2024, TikTok took global action against a covert network linked to **Sputnik Media**. When TikTok removes such networks, it continues monitoring for any attempts to re-establish them. As part of TikTok's anti-recidivism strategy, it removed **11 accounts** in November 2024 that TikTok believed to be associated with Sputnik Media and targeting Romanian and Moldovan audiences.

- **The existence of accounts created solely for the purpose of promoting one presidential candidate;**

6. TikTok understands that this statement ("*The existence of accounts created solely for the purpose of promoting one presidential candidate*") refers to the assertion in Document SRI I that: "*after the creation of the Telegram channel* Redacted *76 new TikTok accounts were registered, having the exclusive role of promoting candidate Călin GEORGESCU*".

7. Given that the ***question 1.1*** does not identify any particular accounts, TikTok is unable to confirm its position in respect of this assertion.

8. TikTok's specialist team identified 76 accounts that violated TikTok's Community Guidelines prohibiting impersonation. TikTok took enforcement action in respect of all 76 accounts.

9. In addition, as referred to in relation to the preceding assertion, TikTok identified and disrupted a CIO network of 78 accounts operating from Romania, with 1,781 followers, that attempted to promote Mr Georgescu on TikTok. To date, this is the only CIO network that TikTok has identified that focussed on promoting Mr Georgescu as a prominent topic. TikTok assessed that this CIO network was operating from Romania and disrupted it in late November 2024. This will be reflected in TikTok's November Covert Influence Operations Report.

---

2 See https://www.tiktok.com/community-guidelines/en/integrity-authenticity.
3 See https://www.tiktok.com/transparency/en-us/countering-influence-operations/.
4 See https://newsroom.tiktok.com/en-eu/continuing-to-protect-the-integrity-of-tiktok-during-romanian-elections.
5 See https://www.tiktok.com/transparency/en-us/covert-influence-operations.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

10. TikTok also has in place as a matter of course scaled detection mechanisms to identify and remove fake engagement, as well as potentially violative automated controlled networks, which can include bots or other spam accounts.

11. Finally, TikTok reiterates that Article 8 and Recital 30 DSA expressly state that the DSA does not impose, and nothing in it should be construed as imposing, a general monitoring obligation or a general active fact-finding obligation on providers, nor a general obligation for providers to take proactive measures in relation to illegal content. To the extent that the Commission identifies any specific accounts, TikTok will immediately review them.

- **The existence of coordinated networks across platforms and other messaging services;**

12. TikTok understands that this statement ("*The existence of coordinated networks across platforms and other messaging services*") refers to the assertion in Document SRI I that: "*The network of accounts directly associated directly with Călin GEORGESCU's campaign initially consisted of 25,000 accounts on the TikTok platform… The activity of the accounts in the network was organized Outside the TikTok platform, namely on the Telegram instant messaging application*".

13. TikTok has limited ability to investigate coordinated networks across other platforms and other messaging services. With respect to the alleged network of 25,000 accounts on the TikTok Platform associated with Mr Georgescu's campaign, TikTok has neither found, nor been presented with, any evidence of such a coordinated network.

14. Likewise, at this point, TikTok has not found any conclusive evidence that coordinated networks originating from another platform or messaging service established a CIO on TikTok to target the Romanian elections.

15. With respect to the Telegram messaging platform, TikTok is continuing to assess and, to the extent relevant, will supplement its response with elements provided in ***Section 1.2*** by next Friday, 13 December 2024.

16. As noted above, in accordance with TikTok's standard processes for enforcing its Integrity & Authenticity Community Guidelines and countering CIOs,[6] TikTok continues to monitor, investigate, and take action against identified attempts to harmfully mislead or deceive its users.

- **The existence of an extensive network of high-profile influencers promoting one presidential candidate on TikTok;**

17. TikTok understands that this statement ("*The existence of an extensive network of high-profile influencers promoting one presidential candidate on TikTok*") refers to the assertion at paragraph 1.3 of Document SRI I that there was: "*an extensive network of high-profile public figures on the TikTok platform (influencers) was used to promote* [Mr Georgescu's] *campaign both directly (by publicly supporting the candidate) and indirectly (through messages that appeared neutral but contained tags associated with the candidate, such as #echilbrușiverticalitate)*".

18. TikTok has investigated a number of instances in which content from various accounts with the hashtag "*#echilbrușiverticalitate*" (the "**Hashtag**") contained similar messaging (not endorsing a particular candidate but suggesting various criteria that the next president should meet). However, to date, TikTok has not identified evidence of a CIO associated with this Hashtag.

19. Finally, TikTok observes that it is usual for online platforms to have hashtags that identify a candidate. The fact that a hashtag identifies a political candidate is not itself an indication of deceptive or manipulative activity.

[6] See https://www.tiktok.com/transparency/en-us/countering-influence-operations/.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

- **The creation of TikTok accounts using the logo and title of Romanian authorities promoting one presidential candidate;**

20. TikTok understands that this statement ("*The creation of TikTok accounts using the logo and title of Romanian authorities promoting one presidential candidate*") refers to the assertion at paragraph 1.4 of Document SRI I that: "[After] *the parliamentary elections of December 1, dozens of TikTok accounts were found falsely using the SRI logo and the Anti-Terrorist Brigade (BAT)... On these accounts, numerous images of Romanian Police forces and BAT, taken from the public space, were distributed*".

21. Given that ***question 1.1*** of ***Section 1.1*** does not identify any particular accounts, TikTok is unable to confirm its position in respect of this assertion. TikTok has not identified such activity in the context of its investigations to detect and disrupt CIOs.

22. TikTok reiterates in this respect that Article 8 and Recital 30 DSA expressly state that the DSA does not impose, and nothing in it should be construed as imposing, a general monitoring obligation or a general active fact-finding obligation on providers, nor a general obligation for providers to take proactive measures in relation to illegal content. To the extent that the Commission identifies any specific accounts, TikTok will immediately review them.

23. TikTok also observes that it has made available to Romanian authorities reporting channels for allegedly illegal content. TikTok expeditiously reviews and, where appropriate, removes access to violative content reported through these channels, generally within 24 hours.

- **The lack of consistent labelling for political accounts;**

24. TikTok understands that this statement ("*The lack of consistent labelling for political accounts*") refers to the assertion in Document SRI I that: "*Posts with purely electoral content from accounts associated with the network promoting Călin Georgescu* [were displayed] *without proper markings and outside the electoral campaign period, aspects prohibited by Romanian electoral legislation*".

25. Under Romanian election law, candidates or other publishers of election propaganda material[7] are required to apply the relevant "*labelling*" or "*markings*" (also referred to as an "**Electoral Code**" or "**CMF Code**") to content that constitutes "**Election Propaganda Material**". TikTok understands the above assertion in Document SRI I to allege that TikTok was obligated to, but did not, remove Election Propaganda Material that depicted Mr Georgescu but did not contain an Electoral Code/CMF Code ("**Unlabelled Georgescu Content**").

26. Without prejudice to whether or not the obligation to remove Unlabelled Georgescu Content is applicable to TikTok, Article 6 DSA makes clear that online platforms shall not be liable for any illegal information that they store unless they have actual knowledge of the specific information at-issue and do not expeditiously remove access. An obligation for TikTok to generally monitor for Election Propaganda Material that should, but does not, include an Electoral Code/CMF Code would be prohibited under Article 8 DSA.

27. In accordance with its obligations under the DSA, TikTok has implemented mechanisms and points of contact through which users, national authorities and others (including trusted flaggers) can report allegedly illegal content. Additionally, TikTok has made such reporting channels available to Romanian national authorities, and generally reviews and takes action in respect of violative content identified through those channels within 24 hours.

28. However, TikTok would also note that the Romanian authorities have sent TikTok certain requests that TikTok considers to exceed the scope of relevant decisions (and that risk unduly restricting the rights of freedom of expression and information in an electoral context), such as: (i) asking TikTok to remove

[7] Which, in broad terms, applies to the publication of content which: (a) directly refers to a candidate or political party participating in the presidential election; and (b) promotes or denigrates those candidates or parties, or encourages people to vote in a certain way.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

all images in relation to Mr Georgescu; and (ii) asking TikTok to take action in respect of accounts which in fact do not appear to have any connection to Mr Georgescu or the Romanian elections.

- **The provider of TikTok concluded in 2021 that a campaign launched in 2020 to promote one presidential candidate constituted a suspicious activity.**

29. TikTok understands that this statement ("*The provider of TikTok concluded in 2021 that a campaign launched in 2020 to promote one presidential candidate constituted a suspicious activity*") refers to the assertion at paragraph 1.5 of Document SRI I that: "*TikTok's first report on the fact that a campaign to promote Călin GEORGESCU is being conducted in the TikTok environment was made in 2020. In 2021 TikTok concluded that this was a suspicious activity*".

30. TikTok does not fully understand this assertion or know its source. TikTok has reviewed internally and not found any basis for this assertion. To the extent that the Commission provides additional information or evidence that substantiates the assertion, TikTok will immediately investigate.

- **The provider of TikTok confirmed that a TikTok user made payments of USD 381.000 between 24 October 2024 and 24 November 2024 to other TikTok' users involved in the promotion of one presidential candidate.**

31. TikTok understands that this statement ("*The provider of TikTok confirmed that a TikTok user made payments of USD 381.000 between 24 October 2024 and 24 November 2024 to other TikTok'* [sic] *users involved in the promotion of one presidential candidate*") refers to the assertion in Document SRI I that TikTok representatives: "*confirmed the identity of the user "bogpr" and mentioned that he made payments in the amount of 381,000 dollars, between 24.10-24.11.2024, to users of TikTok accounts involved in the promotion of candidate Călin GEORGESCU*".

32. TikTok proactively disclosed to ANCOM that a TikTok user (the "**User**") sent TikTok LIVE gifts totalling USD 381,000 to other users between 24 October 2024 and 24 November 2024

33. With respect to the purpose of such payments, TikTok confirmed to ANCOM that it was investigating the User's account to assess allegations that it was involved in CIO which involved offering users payments to promote Mr Georgescu. However, TikTok did not confirm or find evidence that a CIO had taken place, and that remains TikTok's assessment following its investigation.

- **The visibility of one presidential candidate has increased in preference to the other candidates whose posts have been massively filtered, reducing exponentially their presence online.**

34. TikTok understands that this statement ("*The visibility of one presidential candidate has increased in preference to the other candidates whose posts have been massively filtered, reducing exponentially their presence online*") refers to the assertion at paragraph 1.5 of Document SRI I that: "*TikTok did not delete the electoral content as requested by* [the Permanent Electoral Authority], *and it continued to be available to the Romanian public even after the end of the electoral campaign*" and "*As a consequence,* [Mr Georgescu's] *visibility increased preferentially in relation to the other candidates, whose posts were massively filtered, exponentially decreasing their online presence*".

35. Although this assertion is not entirely clear, it appears to suggest that: (i) TikTok allegedly did not remove certain Unlabelled Georgescu Content; and (ii) accordingly, TikTok displayed relatively more Election Propaganda Material that depicted Mr Georgescu than other candidates.

    a. With respect to point (i), please see paragraphs 24-28 above.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

b. With respect to point (ii), although paragraph 1.5 of Document SRI I refers to "*...the other candidates, whose posts were massively filtered, exponentially decreasing their online presence*", Document SRI I does not provide any evidence that other candidates' posts were "*massively filtered*" or "*exponentially decreas*[ed]". It is therefore unclear to TikTok what this assertion is based upon. If any evidence relied upon in reaching this Finding is available, TikTok would welcome the opportunity to investigate the matter further.

36. For the avoidance of doubt, TikTok's recommender systems do not treat any political candidates preferentially.

These assertions may be relevant for the Commission's assessment of the provider of TikTok's compliance with Articles 34 and 35 of Regulation (EU) 2022/2065 in relation to the management of systemic risks of negative effects on civic discourse and electoral processes. Please provide the provider of TikTok's response on each of the assertions made in the intelligence documents declassified by the Romanian President on 4 December 2024 ([7]), indicating, in particular, whether the provider of TikTok acknowledges these findings and/ or supports the claims.

37. TikTok understands that, in this ***question 1*** of ***Section 1.1***, the Commission is asking specifically about the nine assertions listed in the Commission's question.

38. Please see TikTok's responses to each of these assertions in the paragraphs above.

(2) Please explain any changes the provider of TikTok has implemented to its processes, controls, and systems for the monitoring and detection of any systemic risks of negative effects on civic discourse and electoral processes in light of the claims made available by civil society organisations ([8]), Romanian authorities and the intelligence documents declassified by the Romanian President on 4 December 2024 ([9]).

39. In this ***question 2*** of ***Section 1.1***, the Commission refers to "*claims made available by civil society organisations...*", with an accompanying footnote listing examples of such organisations ("*Such as Expert Forum, BROD, EDMO, Alliance4Europe, and Democracy Reporting International*"). While TikTok is aware of certain instances in which certain of these organisations have reported on the Romanian election (and use of the TikTok Platform in connection with the same), it is unclear to TikTok which specific claims made the Commission is referring to.

40. Without prejudice to this point, please note that TikTok has implemented – in addition to its existing measures to mitigate the systemic risk of actual or foreseeable negative effects on civic discourse and electoral processes – a number of changes regarding its monitoring and detection of such risks, which are relevant in the context of the Romanian elections and the claims and documents referred to in the Commission's question. These changes are detailed further in the Newsroom Post, and are summarised below.

41. TikTok recognises that the Romanian elections are a significant event under intense public scrutiny. In this regard, TikTok has been further increasing the measures it has in place to ensure the integrity of discussions on the Platform, including the following:

a. TikTok has strengthened its existing team of over 6,000 Trust and Safety ("**TnS**") professionals dedicated to keeping the Platform safe, adding experts to its Romanian elections taskforce across TnS, including in its Deceptive Behaviour, Security, Ads and other key teams.

b. TikTok has integrated the most recent insights from its expert partners into its policies and guidelines on misinformation and impersonation.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013799

c. TikTok has partnered with the local Romanian NGO 'Funky Citizens' to help TikTok's users safely navigate the Platform and protect themselves against potential misinformation during the Romanian elections.

d. TikTok introduced several new features for the Romanian elections designed to support its users. Specifically, TikTok developed a specific Election Centre for the Romanian elections, and introduced a Short Video Notice Tag. The in-app Election Centre connects TikTok's users to trusted official information, guiding them to the Permanent Electoral Authority's page. Furthermore, the Election Centre provides local users with access to videos created in collaboration with Funky Citizens to help them critically analyse online election-related content.

42. In countering CIOs during the Romanian elections, TikTok proactively seeks and continuously investigates leads for potential CIOs. TikTok's proactive measures include looking for signs of coordinated inauthentic behaviour and carefully considering activities and context on other platforms to identify and prevent CIOs and deceptive behaviours on the Platform.

43. As the Commission is aware, TikTok is engaging with Romanian authorities and has encouraged such authorities to share any relevant intelligence, so that TikTok can work together with them to help ensure the integrity of the Romanian elections. The CIO networks that TikTok has detected specifically targeting the Romanian elections (see paragraph 5 above) have so far been relatively small-scale, domestic operations coordinated on TikTok.

44. Since 2020, TikTok has actively participated in the EU Code of Practice on Disinformation's ("**COPD**") Taskforce and working groups, including co-chairing the Elections subgroup and Transparency subgroup, in order to stay at the forefront of combating election misinformation – including through collaboration with the Commission and signatories on trends and mitigation strategies. Some of the NGOs referred to in ***footnote 8*** of the RFI are also COPD signatories, and TikTok therefore engages with them in that context. Further, BROD (the Bulgarian-Romanian European Digital Media Observatory-related hub) and Democracy Reporting International are two NGOs which have been sending TikTok reports through the Rapid Response System established under the COPD.

Exhibit 267

TikTok Report - 12 December, 2024

Overview

Reports from the Romanian authorities have practically eased off, however, our communication channels are open and our teams continue to monitor for any violating activity on the platform.

Incidents and Actions

- We have identified and removed some further impersonation accounts. Since our last communication, we have identified and removed over 280 accounts impersonating Romanian Government, Politician, or Political Party Accounts (GPPPAs).
- A temporary search restriction for the keywords *'lovitura de stat'* (coup d'etat) and *'razboi'* (war) has been imposed in Romania for the next 60 days to mitigate the risk of harmful narratives.
- The teams have identified and enforced content related to the following trends/claims:
    - *Misinformation related to Ukrainian children in Romania receive significantly higher allowances than Romanian children;*
    - *Video and audio content containing sexual harassment towards the Romanian Prime Minister;*
    - *Audio used for degrading statements targeting politicians, regular users, and the LGBT community.*

  None of this content has been in large volumes.

- As part of our engagement with the Rapid Response System (RRS) under the Code of Practice on Disinformation, we have followed up in writing to Democracy Reporting and BROD (EDMO) and all current reports have been closed out. The RRS will remain active until Monday December 16th, per agreement with the Commission / Code group.
- No incidents of note on the **Covert Influence Operations ('CIO') front**. We investigated an escalation regarding potential fake engagement linked to 2 accounts, one of which was related to the new TikTok account that was set up by Calin Georgescu over the weekend. In this latter case, the video views and followers were found to be authentic.
- As regards enforcement of **political advertising,** there have been no major updates. We continue to do 'sweeps' but no leakage has been detected.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 268

TikTok Technology Limited's response to RFI
(Ref. Ares(2024)8523013 - 29/11/2024)
13 December 2024

Introduction

1. TikTok Technology Limited submits this response (the "**Response**") to the Commission's Request for Information Ref. Ares(2024)8523013 (the "**RFI**") as the legal entity that is the provider of the designated very large online platform TikTok (the "**Platform**") in the EU under Regulation (EU) 2022/2065 (the "**DSA**").

2. As a general point, TikTok notes that a number of the Commission's questions relate to, e.g., specific items of content, specific orders, or specific reports. TikTok reserves its position with respect to the proportionality and relevance of such questions in the context of assessing TikTok's compliance with Articles 34 and 35 DSA.

3. TikTok has been working diligently on identifying documents responsive to questions where the Commission has requested documents. TikTok wishes to emphasise that identifying, reviewing and preparing documents for production takes time, including from a technical standpoint. Together with this Response, TikTok provides screenshots of the relevant dashboard in response to the first part of ***question 14***. For other questions, TikTok will provide the Commission with relevant documents by 31 December 2024.

Questions and Answers

1. Please provide the orders or notices that the provider of TikTok received, between 1 September 2024 and 24 November 2024, from national authorities in the context of the 2024 Romanian Presidential and Parliamentary elections (hereinafter 'the Romanian elections'), all internal documents produced in the provider of TikTok's assessment of these orders or notices, all documents, including communications, that show the effect given to these orders or notices, including the timing of this effect, if any, and the reply that the provider of TikTok sent in response to these orders or notices.

4. TikTok understands the reference to "*orders or notices*" in this ***question 1*** to mean: (i) orders to act against illegal content received from Romanian national authorities in accordance with Article 9 DSA; and (ii) notices received from Romanian national authorities via TikTok's DSA notice and action mechanism for reporting specific items of illegal content in accordance with Article 16 DSA.[1]

5. Between 1 September 2024 and 24 November 2024 (the "**Relevant Period**"):

   i. TikTok is not aware that it received any Article 16 notices via its dedicated mechanism from Romanian authorities; and

   ii. To the extent that any orders were received from Romanian authorities, these orders (and the manner in which they were communicated to TikTok) did not meet the criteria stipulated in Article 9 DSA.

6. Notwithstanding the above, to assist the Commission's consideration of the circumstances surrounding TikTok's good faith engagement with Romanian national authorities concerning the Romanian elections (i.e. beyond the parameters of Article 9 or Article 16 DSA), TikTok hereby provides information in relation to relevant communications that it had with such authorities during the Relevant Period whereby Romanian authorities notified or requested the removal of specific items of allegedly illegal content and/or transmitted to TikTok (general) orders that referred to legal grounds. At the outset, TikTok wishes to note that its approach to such engagement has been carried out in good faith, which the DSA expressly states "*should include acting in an objective, non-discriminatory and proportionate manner, with due regard to*

[1] TikTok further notes that there is no obligation under the DSA to comply with a request from a national authority to action specific items of illegal content, even if a request to do so is received under Articles 9 or 16 DSA.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

the rights and legitimate interests of all parties involved, and providing the necessary safeguards against unjustified removal of legal content…".[2]

7. On this basis, TikTok's communications with Romanian national authorities during the Relevant Period concerning the Romanian elections can be generally summarised as follows: (i) communications from Romania's Permanent Electoral Authority ("**AEP**") ("**Category 1**"); and (ii) communications from the Romanian National Police ("**Category 2**").

8. In respect of Category 1, TikTok notes as follows:

   i. Prior to 21 November 2024, TikTok received a small number of communications from the AEP containing informal requests. Each of these communications was sent by informal email to TikTok's Government Relations & Public Policy Manager ("**TikTok GR/PP**") for Romania. The emails did not purport to be or pertain to formal orders or authorisations; they simply contained informal requests set out in the body of an email. For example, one of these requests was for TikTok to remove content on the basis that it was "*disrespectful and insults the PSD party*", but without stating any specific legal basis in support of this request. TikTok twice sought further details of the legal basis for this request, but the AEP provided no such details in response.[3] In a subsequent email, the AEP reported a suspected 'deepfake' in "*a video montage meant to discredit a presidential candidate*". The reported video content was assessed by TikTok's Trust & Safety team and removed (within approximately 24 hours) for violating TikTok's Community Guidelines (the "**Community Guidelines**").

   ii. From 21 November 2024, the AEP began to attach decisions of the Romanian electoral authorities to certain of their email requests (i.e. decisions from either the Central Electoral Office or County Electoral Offices). Each of these decisions were general in nature and were not specifically addressed to TikTok. In terms of the approach for the communications attaching these decisions, the AEP, for example, set out in its emails (again in a highly informal manner) URLs to content that the AEP alleged was illegal based on the particular decision attached. TikTok notes that several of the communications from the AEP sought to convey a very broad interpretation of the relevant decision attached. It is also important to note that none of these communications met the criteria specified under Article 9(2) of the DSA. For example, because they did not clearly: set out their legal basis; contain a statement of reasons explaining why the information is illegal content; contain information about redress mechanisms; and/or specify their territorial scope. In addition, they did not contain an English translation of the core elements and they were not transmitted to TikTok's single point of contact ("**SPOC**"), as required by Article 9(2)(c) DSA.

   iii. In light of the above, in handling these communications in good faith, TikTok's teams were wary of the very informal approach adopted by the AEP in the context of the Romanian elections, in particular the potential for political influence on the process and/or the unjustified removal of legal content (such as political speech). In these circumstances, and given that none of the communications met the requirements stipulated under Article 9 DSA (notwithstanding several requests from TikTok in this regard) – TikTok nonetheless acted in good faith and assessed content identified in these requests: (i) against TikTok's Community Guidelines; and/or (ii) for compliance with Romanian electoral laws (based on advice from external Romanian legal counsel).

9. In respect of Category 2, TikTok notes as follows:

   i. TikTok received two relevant requests from the Central Intelligence Analysis Unit of the Romanian National Police. These requests were informal in nature and were each sent by email to TikTok's Law Enforcement Outreach team (whose function it is to engage with law enforcement authorities). Both requests, sent on 21 and 23 November 2024, respectively, contained URLs to specific content on the Platform that the Romanian Police alleged may be

[2] Recital 26 DSA.

[3] Indeed, TikTok subsequently received a communication relating to similar content from Romania's Minister for Research, Innovation and Digitalisation.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

illegal content under Romanian electoral law (with particular regard to election silence period). As above, these were informal requests, which also did not meet the criteria specified under Article 9 DSA.

ii. As above, given that these were informal requests, TikTok nonetheless acted in good faith and assessed the content referenced in these requests: (i) against the Community Guidelines; and/or (ii) for compliance with Romanian electoral laws (based on advice from external Romanian legal counsel).

10. Finally, TikTok received two letters from Romania's National Authority for Management and Regulation in Communications ("**ANCOM**") during the Relevant Period. Both letters were general in nature and did not notify or request the removal of specific allegedly illegal content under Romanian law.

11. TikTok hopes that the foregoing addresses the Commission's concerns and demonstrates that TikTok diligently and expeditiously engaged with requests from Romanian authorities, above and beyond its specific obligations under Articles 9 and 16 DSA, in an attempt to consider their communications in a good faith and diligent manner, consistent with its policies and its aim of mitigating risk of undue political influence on its Platform, while also applying appropriate safeguards against unjustified removal of legal content.

12. Please see further TikTok's response to ***question 2*** below regarding communications discussed in paragraphs 7 and 8 above.

2. For each order or notice received by the provider of TikTok between 1 September 2024 and 24 November 2024 in the context of the Romanian elections, please provide the time taken to inform the authority which issued the order or notice of its receipt and the effect given to it, and the time to give effect to it.

13. As explained above in response to ***question 1***, TikTok did not receive from Romanian national authorities during the Relevant Period any orders that met the requirements of Article 9 DSA. Accordingly, no specific obligations were triggered under that Article to, e.g., inform such authorities of the effect given to any such orders.

14. Notwithstanding this, TikTok provides the following information regarding TikTok's good faith engagement with Romanian authorities in the context of the Romanian elections, whereby those authorities notified or requested the removal of specific items of allegedly illegal content under Romanian law:

Sender	Recipient	Subject matter	Date received	Date of response(s)	Effect given to request[4]	Notes
Category 1:						
AEP	TikTok's GR PP for Romania	Emails from AEP in respect of specific videos / audio identified by URL and allegedly insulting to the PSD Party	14 November 2024 (1:51pm and 3:14pm GMT)	Email response from TikTok GR/PP for Romania on 14 November 2024 (at 3:09pm and 4:53pm GMT) seeking further details of the legal basis for this request	N/A – as the identified content did not violate TikTok policies, and the AEP did not respond to TikTok's request for the legal basis	
AEP	TikTok GR/PP for	Email from AEP in	15 November	Email responses	Our records	

[4] Please note that the relevant system indicates times in GMT-1 time-zone.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013639

	Romania	respect of alleged deepfake of presidential candidate identified by URL	2024 (8:19am GMT)	from TikTok GR/PP for Romania on 15 November 2024 (at 8:42am GMT), confirming receipt and that the video has separately been reported	indicate that the identified video was removed for violating TikTok's Community Guidelines on 16 November 2024 (by 10:43am GMT-1 / 11:43am GMT)	
AEP	TikTok GR/PP for Romania	Emails from AEP in respect of Decision Nr. 175D/ 20.11.2024 of Biroul Electoral Central requesting that TikTok "*remove all online materials of candidate Calin Georgescu that do not contain the financial representative identification code*", and initially listing 2 specific account URLs. The subsequent email listed 14 video URLs "*on the candidate account*", and a further 6 video URLs relating to "*supporters' accounts*"	21 November 2024 (7:30am and 7:58pm GMT)	Email responses from TikTok GR/PP for Romania on: 21 November 2024 (at 8:40am GMT) confirming receipt; and 22 November 2024 (at 11:47am GMT) advising re: action taken	Our records indicate that access to 14 videos was blocked in Romania on 22 November 2024 (by 9:56am GMT-1 / 9:56am GMT) for violating local law (per the relevant decision); a further 3 videos were removed for violating TikTok's Community Guidelines on 25 November 2024, and a further video on 29 November 2024	Further email responses from TikTok GR/PP for Romania on 25 November (at both 5:22pm and 5:47pm GMT) and on 29 November 2024 (at 12:39am GMT), again advising re: action taken
AEP	TikTok GR/PP for Romania	Email from AEP in respect of Decision Nr. 54 / DIN 21.11.2024 of Birou; Electoral de Circumscripţie Judeţeană Nr. 3 Arges requesting that TikTok "*eliminate the materials as illegal*" relating to "*electoral campaign materials…*", relating to a specified account, and listing 4 specific video URLs	21 November 2024 (7:16pm GMT)	Email response from TikTok GR/PP for Romaniat on 22 November 2024 (at 11:47am GMT) advising re: action taken	Our records indicate that access to 4 photo posts was blocked in Romania on 22 November 2024 (by 8:56am GMT-1 / 9:56am GMT), for violating local law (per the relevant decision)	
AEP	TikTok GR/PP for Romania	Email from AEP in respect of Decision Nr. 39 / 22.11.2024 of Biroul Electoral de Circumscripţie Cetăţenii Români cu Domiciliul Sau Reşedinţa în Afara Ţării no. 43 requesting that TikTok remove "*a sound background […] with indecent, vulgar, trivial connotations addressed to the Social Democratic Party, as well as the PSD President Ion-Marcel*	22 November 2024 (6:37pm GMT)	Email response from TikTok GR/PP for Romania on 29 November 2024 (at 12:39pm GMT) advising re: action taken	Our records indicate that access to the reported audio was blocked in Romania on 23 November 2024 (by 8:37pm GMT-1 / 9:37pm GMT), for violating local law (per the relevant decision)	A similar communication in relation to Decision Nr. 39 was also received from the Ministry of Research, Innovation and Digitalisation on 23 November 2024

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013640

		Ciolacu..."				
Category 2:						
Romanian National Police	TikTok Law Enforcement Outreach team point of contact	Email from Romanian National Police requesting assistance in relation to compliance with Romanian electoral laws, and citing specific content	21 November 2024 (at 8:09pm GMT)	Email response from TikTok Law Enforcement Outream team point of contact on 22 November 2024 (at 10:26am GMT), confirming that the reported content would be reviewed	Our records indicate that access to 20 video URLs were blocked in Romania on 23 November 2024 (by 3:49pm GMT-1 / 4:49 GMT), for violating local law	This communication involved sharing of "sample" content that may be in violation of election silence period rules on 21 November 2024 (in advance of that period commencing on 23 November 2024)
Romanian National Police	TikTok Law Enforcement Outreach team point of contact	Email from Romanian National Police requesting assistance in relation to compliance with Romanian electoral laws, and citing specific content	23 November 2024 (at 6:41am and at 1:22pm GMT)	Email response from TikTok Law Enforcement Outream team point of contact on 23 November 2024 (at 11:45am and at 2:02pm GMT), confirming that the reported content would be reviewed	As above.	

3. In its Guidelines on the mitigation of systemic risks for electoral processes, the *"Commission recommends that providers of VLOPs and VLOSEs regularly and, where necessary, urgently, exchange information with and have contact points for competent national and European authorities, involving the Digital Services Coordinator of their Member State and, where appropriate, competent regional and local authorities, to facilitate the exchange of information, as well as the Commission, as appropriate".* Please explain whether the provider of TikTok has put in place any specific contact points for such purpose in the Romanian elections context and what contacts it has had with Romanian authorities. Please include a list of the authorities the provider of TikTok is in contact with.

15. As a preliminary point, TikTok notes that Communication C/2024/2537 (the "**Elections Guidelines**") contains non-binding recommendations for "*possible measures*". Without prejudice to this, TikTok provides the following response to this question.

16. At the outset, TikTok understands the language from the Elections Guidelines quoted in this ***question 3*** as referring to a recommendation that competent national authorities and VLOP providers exchange contact information to enable competent authorities to share information that can: (i) inform VLOPs' risk assessments and mitigation measures on electoral processes; or (ii) inform national authorities' actions that are within their competence to protect the integrity of electoral processes (see paragraphs 44-45 of the Elections Guidelines). As explained in response to ***questions 1*** and ***2*** above, Romanian authorities were in direct contact with TikTok concerning the Romanian elections.

17. Specifically, TikTok established the following specific contact points to facilitate the exchange of information with Romanian authorities in the context of the Romanian elections:

    i. Contact information for TikTok's GR/PP manager for Romania was exchanged with relevant Romanian authorities for the purposes of TikTok's voluntary adoption of the foregoing recommendation. The AEP agreed to this arrangement on 22 October 2024 and ANCOM subsequently acknowledged this arrangement via email on 31 October when sending TikTok a letter requesting us to analyse a request from the AEP.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013641

ii. Contact information for TikTok's Law Enforcement Outreach team was exchanged with Romanian law enforcement authorities, in particular the Romanian National Police.[5]

iii. In addition, TikTok has also made available its dedicated Safety Enforcement Tool[6] ("**TSET**") to Romanian authorities – including to the Romanian National Cyber Security Directorate ("**DNSC**") and the Romanian Ministry of Research, Innovation and Digitalisation ("**MCID**") – to further facilitate streamlined and efficient exchange of information. Further, the AEP has been onboarded to T-SET in the context of the Romanian elections since 29 November 2024.

18. The Romanian authorities with which TikTok has been in contact include:

i. the AEP;
ii. ANCOM;
iii. the Romanian National Police;
iv. MCID;
v. DNSC;
vi. the Romanian Ministry of Foreign Affairs ("**MAE**");
vii. the Romanian National Audiovisual Council ("**CNA**"); and
viii. the Romanian Intelligence Service ("**SRI**").

4. Please explain whether the provider of TikTok has put in place any specific measures to reinforce its internal processes, resources, testing, documentation, or supervision of the service in particular as regards detection and mitigation of systemic risks related to civic discourse and electoral processes in the context of the Romanian elections. If yes, please detail these measures.

19. TikTok understands this question as relating specifically to any additional measures taken in the context of the Romanian elections besides those taken in the ordinary course of business during an election through TikTok's Elections Integrity Programme.

20. Concerning TikTok's work in relation to the Romanian elections, please see the newsroom post published by TikTok at 7:30 p.m. CET on 6 December 2024 (the "**Newsroom Post**"),[7] and the response to ***question 2*** to ***Section 1.1*** of the Commission's Request for Information Ref.Ares(2024)8716531 submitted on 7 December 2024.

21. In connection with TikTok's Election Integrity Programme, TikTok sets out below an overview of the key mitigations and processes implemented in the context of the Romanian elections. As noted in TikTok's 2023 Systemic Risk Assessment, TikTok's preparation for an election starts many months before the election, and begins by way of risk assessments of the political landscape facilitated via external engagements and consultations with local experts. For Romania, the risk assessment and related scenario-planning conducted through TikTok's Elections Integrity Programme commenced on 13 May 2024.

22. **Key processes**: In addition to the risk assessment and scenario-planning processes (described above), the following processes are important parts of the Programme to ensure effective measures are in place for specific elections:

i. **Partner onboarding**: Separate from its engagement with reporters designated as trusted flaggers under the DSA, TikTok operates a Community Partner Channel ("**CPC**") whereby

---

5 In the run up to the Romanian presidential elections, TikTok's Law Enforcement Outreach team had been in ongoing communication with the Romanian National Police. With specific regard to the Romanian elections, the Law Enforcement Outreach team was in contact by email with the Romanian National Police on 31 October 2024 and held a virtual meeting on 8 November 2024, which covered the upcoming elections (among other topics).

6 See https://safety-enforcement.tiktok.com/.

7 See https://newsroom.tiktok.com/en-eu/continuing-to-protect-the-integrity-of-tiktok-during-romanian-elections.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

onboarded NGOs perform a similar role to trusted flaggers to submit reports of suspected harmful content, including Election Misinformation.

- TikTok has granted 3 Romanian organisations (Ora de Net Hotline, Active Watch, and Save the Children Romania) access to its CPC. The CPC is an escalation channel using TikTok Safety Enforcement Tool (TSET) allowing the organisations to escalate any content directly to TikTok's Incident Management team for review.

ii. **Fact-checking**: A key component of election integrity preparations is assessing the need to leverage partnerships with professional third-party fact-checkers to assist TikTok with addressing misinformation. All fact-checking partners must be signatories of the International Fact-Checking Network ("**IFCN**"), a global network and governing body of professional fact-checkers that annually reviews members to ensure they uphold the IFCN Code of Principles including: non-partisanship, fairness, transparency, and an honest corrections policy.

- In Romania, TikTok partners with LeadStories, which is a verified member of IFCN and the European Fact-Checking Standards Network (EFCSN).[8]

iii. **Operations moderation resources (capacity planning)**: Prior to launching the Fact-Checking Program in a specific market, TikTok stands up a specialised moderation team to manage the fact-checking escalation queues and receive specialised integrity and authenticity training to up-skill their policy knowledge.

- As noted in the Newsroom Post, TikTok has 95 Romanian-speaking moderators, which is the largest such team among digital platforms in the country, both in absolute terms and relative to the number of users.

iv. **Mission control centre**: For elections, TikTok will convene a "Mission Control Centre" to have centralised coordination and a dedicated and highly engaged XFN group focused on potential election related issues in the run up to an election, while also allowing for efficient, rapid, real-time, cross functional alignment and decision-making.

- Establishing the "Mission Control Centre" is an important process to coordinate the election strategy. The Mission Control Centre for the Romanian election was established on 23 October 2024 (before the official campaign period commenced).

23. **Front-end measures**: Depending on the election, some or all of the following front-end measures may be implemented, which are visible to TikTok users:

i. **Elections Centre**: This is an in-app election information center that provides users with accurate up-to-date information about an election.

- On 11 November 2024, TikTok launched a dedicated in-app Election Centre to provide Romanian users with reliable and authoritative information as the country prepared to go to the polls. TikTok partnered with Funky Citizens, a local fact-checking organisation, to provide media literacy tips alongside key election and voting information from the Electoral Commission. When on TikTok, users were directed to the Election Centre through prompts and labels on relevant election content and when making election-related searches. TikTok took a similar approach during the EU Elections earlier this year, connecting hundreds of thousands of users to TikTok's EU Election Centre. Please see TikTok's Newsroom post on EU elections: https://newsroom.tiktok.com/en-eu/how-we-kept-our-community-safe-during-the-2024-european-parliament-elections.

ii. **Search banners**: This is a banner that is displayed in the Search section on the TikTok app. The banner appears at the top level search result for a given set of keywords, there is a pool of 217

[8] See https://efcsn.com/verified-members.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013643

keywords for the Romanian Election. The banner is clickable and serves as one of the entry points to the Election Centre.

- In Romania, TikTok launched this feature on 11 November 2024, redirecting users to the dedicated Election Centre.

iii. **Video notice tags**: A general election notice tag is an information bar at the bottom of a video which is automatically applied to over 50 specific hashtags (or set of hashtags) for the Romanian Elections. The information bar is clickable and will direct users to an in-app guide (or reliable third party resource). This feature can be utilised as an entry point (among others) for an Elections Centre or point to authoritative external resources such as a country's Electoral Commission. TikTok adopts a similar approach for LIVE videos.

- In Romania, TikTok launched this feature on 11 November 2024, redirecting users to the dedicated Election Centre.

iv. **Unverified Content Label**: In a dynamic election situation where events are unfolding, TikTok can add an "unverified content label". This is a misinformation product feature aimed at reducing engagement and sharing of content referring to unfolding events, which is active in those EU countries where the Fact-checking Programme is rolled out, including in Romania.

v. **Election Misinformation Reporting**: This feature enables users to report potential election misinformation through a specific election report reason, such as "Election Misinformation" across different reporting mechanisms.

24. **Back-end measures**:

i. **Measures to detect and disrupt covert influence operations**: TikTok has provided details regarding its on- and off-Platform detection of, and actions taken against, covert influence operations ("**CIOs**") in the context of the Romanian elections, and the results of those actions, in its responses to ***question 1*** of ***Section 1.1*** and ***question 6*** of ***Section 1.2*** of RFI Ref. Ares(2024)8716531 (the "**Second Romanian Elections RFI**"). Please also see TikTok's responses to ***questions 5(c)*** and ***11(b)*** of RFI Ref.Ares(2024)6974764 (the "**Recommender Systems & Elections RFI**"), in which TikTok provides details regarding its activities to detect and disrupt CIOs in general.

ii. **Real-time election monitoring**: TikTok's election teams engage in real-time monitoring using centralised Election dashboards. The real time Election dashboard for a particular election is an important source of information on evolving trends/threats on the Platform, while also providing an overview and sample of the types and volumes of content relevant to an election. The dashboard helps TikTok's elections team identify potential violative content and take informed decisions, including to add new violating keywords or hashtags to keyword lists.

- As mentioned in response to ***question 23.iv***, TikTok established a Mission Control Centre for the monitoring of the Romanian election on 23 October 2024.

iii. **Keyword strategies**: Use of keyword strategies to detect and action potential Election Misinformation is a core part of the Programme. This involves the development of lists of keywords associated with an election or with Election Misinformation (and which may include emojis and special characters). In the context of election integrity, where those keywords are contained in video titles or hashtags, TikTok utilises word matching to detect and recall potentially violating content for review, or to limit content visibility. During an election period, TikTok may supplement the list with keywords related to voter suppression, voter fraud, and election disinformation. TikTok's policy team reviews cases surfaced by the dashboard, and requests keyword tracking additions and updates based on violative trends that may arise during the electoral period.

- Please see TikTok's response to ***question 12***.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013644

iv. **Sweeps**: TikTok's policy team also conducts sweeps of election-related content in the run-up to elections as a risk mitigation measure to detect risks and understand risk trends. For example, this may involve conducting an assessment of the top reported videos associated with the election.

- Please see TikTok's response to ***questions 12*** and ***13***.

v. **Intake signals/reports**: Reports of potential Election Misinformation and other election-related intelligence is obtained from a range of sources, including:

- Threat intelligence partnerships: TikTok works with external threat intelligence partners to provide coverage for ongoing election integrity research and intelligence. Please note that a number of intelligence functions address election related risk at TikTok. The Influence Operation Intelligence and Discovery ('IOID') function focuses on identifying covert influence operations operating on TikTok, while the Converged Risk Intelligence function brings together diverse risk domains such as physical security, geopolitics, technology and regulatory environments. By converging these domains, the intelligence team ensures a holistic understanding of threats, recognizing that risks in one area often intersect with or amplify in another. These intelligence functions leverage open-source research while respecting best industry practices and regulations, as well as information from private external intelligence providers to assess and mitigate risks affecting the platform and its users. The accuracy of the findings of the external research is then verified through an analysis on the platform using internal tools and systems. The finalized intelligence briefs inform the Influence Operation Intelligence and Discovery team in charge of CIO and policy teams on risk hypotheses related to elections and other critical areas. Once the Influence Operation Intelligence and Discovery team receives the reports they treat them as high confidence hypotheses but not as finalized investigations. They use the reports to conduct their own investigation. The policy team then enforces the conclusion of their investigation.

- Risk Analysis: The Trust & Safety Risk Analysis function provides information on election-related misinformation/conspiracy theory trends across other platforms, allowing us to proactively search for similar content on TikTok with a view to pre-empting escalations. The team also details specific trends for the triage team to assign for further investigation.

- Public Policy Reporting: The Government Relations Team provides an escalation point for appropriate external sources to flag potential mis/disinformation. Emails received are routed to a triage team to discuss an appropriate enforcement and response.

  - Please see TikTok's responses to ***questions 1***, ***2***, and ***5*** of the RFI regarding TikTok's engagement with external sources in the context of the Romanian elections.

- Internal Escalations: Both the investigations and market analysts teams conduct proactive on-platform investigations to identify mis/disinformation cases and trends. Suspected cases are sent directly to the triage team.

  - Please see response to ***question 8(d)*** regarding TikTok's proactive measures in the context of the Romanian elections.

- Government, Politician, and Political Party Accounts ("**GPPPAs**"): For accounts falling within TikTok's GPPPAs definition, TikTok applies a label in its system. See further, e.g., TikTok's responses to ***question 5(a)*** of the Recommender Systems & Elections RFI and ***question 9(c)*** below.

25. **External engagement, communication and education measures**:

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

i. **Engagement with reputable Election Commissions/Authorities**: Through TikTok's Government Relations and Outreach and Partnership Management ("**OPM**") teams, TikTok engages with reputable election commissions/authorities to obtain authoritative information regarding the conduct of elections under their remit.

ii. **Campaign teams**: Through local Government Relations teams, TikTok also establishes direct communication channels with the campaign teams of candidates/political parties with a view to ensuring prompt action in relation to emerging and evolving issues (and to pre-empt and explain avoidable issues, such a candidate's content being inadvertently removed for copyright violations).

TikTok engaged with the main political groups to inform them about TikTok's policies and educate political actors about TikTok's safety measures:

- 29 October 2024, with the Social Democratic Party ("**PSD**");
- 4 November 2024, with the Alliance for the Union of Romanians ("**AUR**");
- 5 November 2024, with the Renewing Romania's European Project Party ("**REPER**");
- 5 November 2024, with the Save Romania Union Party ("**USR**");
- 7 November 2024, with the Democratic Alliance of Hungarians in Romania ("**UDMR**"); and
- 7 November 2024, with the United Right Alliance ("**FD**" and "**PMP**").

iii. **External Election Speaker Series**: As part of the Elections Integrity Programme, TikTok's OPM team undertakes various initiatives, including an External Election Speaker Series. This involves OPM coordinating an elections speaker series in the lead-up to elections with suitably qualified external local/regional experts who share their insights and market expertise with TikTok's internal cross-functional teams on key issues involved in and impacting the election and potential risks that may impact or manifest on the Platform.

- Please see TikTok's response to ***question 5***.

iv. **Media Literacy**: Before elections to help users, TikTok partners with reputable NGOs and other partners to make available resources to educate users on how to detect and assess potential Election Misinformation.

- TikTok's media literacy campaign on the Romanian elections, developed together with Funky Citizens, reached 29.5 million views. Additionally, please see TikTok's response to ***question 5*** regarding TikTok's engagement with NGOs and other partners in the context of the Romanian elections.

5. If applicable, please describe if and how the provider of TikTok has involved representatives of the recipients of the service, representatives of groups potentially impacted by their services, independent experts and civil society organisations in the design of the measures referred to in question 4.

26. TikTok engages with elections experts in relation to elections in multiple ways. This is done primarily through engagement by TikTok's Government Relations & Public Policy teams and OPM teams, as well as via TikTok's Trust and Safety, Legal.

27. To ensure the teams are as prepared as possible for upcoming elections TikTok invites its external fact-checking partners to attend a "Speaker Series" session with its internal teams. During this session, fact-checking experts present key potential risks and challenges during the election period directly to the TikTok's Trust and Safety teams responsible for monitoring the election. These sessions play a key role in

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TikTok's internal election-related preparations, by ensuring that all relevant internal teams are abreast of ongoing risks during the election period. LeadStories serves as the fact-checking partner for Romania, and the platform continues to expand its fact-checking network.

28. For the Romanian elections specifically, the session was conducted with TikTok's Media Literacy Partner in Romania, Funky Citizens. This is a continuation of TikTok's collaboration with Funky Citizens, which previously provided a Speaker Series for TikTok in May 2024 with insights relating to the EU elections. TikTok has also collaborated with Funky Citizens to produce media literacy content in Romanian, giving users tips to combat misinformation and disinformation. The Romanian community can find the video series with practical advice and useful information about the electoral process on Funky Citizens' official TikTok account, and in the in-app Election Center dedicated to Romania's elections.

29. In addition, with respect to the Romanian elections, TikTok organised or participated in the following sessions and calls:

30. An online meeting with Funky Citizens was held on 11 October 2024, to agree on a joint media literacy campaign, which was initiated on 22 November 2024.

31. Between 29 October and 7 November 2024 TikTok's Government Relations team organised dedicated online sessions with political groups in Romania, to inform them about TikTok's elections-related policies. This included educating key political entities and persons about safety measures. These sessions took place on:

    i. 29 October 2024, with the Social Democratic Party ("**PSD**");
    ii. 4 November 2024, with the Alliance for the Union of Romanians ("**AUR**");
    iii. 5 November 2024, with the Renewing Romania's European Project Party ("**REPER**");
    iv. 5 November 2024, with the Save Romania Union Party ("**USR**");
    v. 7 November 2024, with the Democratic Alliance of Hungarians in Romania ("**UDMR**"); and
    vi. 7 November 2024, with the United Right Alliance ("**FD**" and "**PMP**").

32. TikTok's user empowerment tools, such as search interventions or Election Centre, are informed by the information gathered, but not limited to, through the engagement with elections experts and representatives.

6. If applicable, please submit the correspondence with the above stakeholders, as well as any minutes of meetings conducted with these stakeholders, in relation to the preparations for the Romanian elections.

33. Please see paragraph 3 above.

7. On the basis of a potential analysis of the Romanian political and legal context related to elections, please clarify whether the provider of TikTok has put in place any specific measures vis-à-vis national and local rules and regulations.

34. First, as the Commission is aware:

    i. Article 8 and Recital 30 DSA expressly state that the DSA does not impose, and nothing in it should be construed as imposing, a general monitoring obligation or a general active fact-finding obligation on providers, nor a general obligation for providers to take proactive measures in relation to illegal content; and

    ii. Article 14(4) DSA requires providers to enforce restrictions in their terms and conditions in a "*diligent, objective and proportionate manner*" with due regard to, among other things, users' fundamental rights such as freedom of expression.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TikTok understands that nothing in this RFI should be construed as suggesting otherwise.

35. With regard to the political context relating to the Romanian elections, TikTok's Government Relations & Public Policy teams have closely monitored the political landscape, including through reporting from external agencies (covering topics such as: the election timeline, recent EU parliament results, key takeaways from past elections, the presidential race, the profiles and policies of the leading presidential candidates, the parliamentary race). In advance of the Romanian elections, TikTok's Government Relations & Public Policy teams also undertook analysis to inform a media literacy campaign to counteract disinformation in the context of the elections.

36. With regard to the legal context relating to the Romanian elections, TikTok has developed Moderator Guidelines (the "**Moderator Guidelines**") in relation to elections content for EU countries for the purposes of processing requests relating to illegal content under Article 16 DSA. These Moderator Guidelines were developed before the Romanian elections and specifically cover topics including election silence periods, the sharing of opinion polls, ballot papers and support for candidates. The Moderator Guidelines are jurisdiction-specific and were developed with regard to national rules and regulations.

37. In addition, TikTok has ensured that it has access to specialist external local legal counsel in Romania to advise on specific questions of Romanian law that may arise in relation to the Romanian elections. In addition to the local expertise within TikTok's specialist teams (e.g., TikTok's local Government Relations team), these external legal counsel have an understanding of the relevant political context in Romania in connection with which they advise.

8. According to TikTok's advertising policies ([9]), TikTok does not allow "*ads featuring political content across any of [its] monetization features, including paid ads, creators being paid to make branded political content, and other promotional tools on the platform*".

a. Please provide a detailed description of how the provider of TikTok defines "political content" as well as any internal guidelines for identifying such content.

38. In line with TikTok's obligation to consider users' fundamental rights and freedom of expression, TikTok users are allowed to share content relating to political themes or issues provided that it does not constitute an advertisement or branded content (as explained below) and complies with applicable policies, including TikTok's Community Guidelines.

39. TikTok does not allow political content to feature in advertising (i.e., content that TikTok is remunerated specifically to promote, "**Ad(s)**") or in branded content (e.g. a user promoting content for a third party in exchange for payment or any other incentive, where TikTok is not remunerated to promote that content, "**Branded Content**").

40. In this context, "*political content*" means the types of content relating to political themes or issues that are prohibited by: (i) TikTok's Politics, Culture, and Religion Ads Policy[10] (the "**Political Ads Policy**"); and/or (ii) TikTok's Branded Content Policy[11] (the "**Branded Content Policy**").

41. To assist in enforcing the Political Ads Policy, TikTok classifies certain political accounts as "*Government, Politician, and Political Party Accounts*" ("**GPPPAs**") and applies restrictions to such accounts, including disabling their access to advertising features to prevent them from creating Ads (as noted above, as a limited exception, TikTok allows governments to advertise in limited circumstances such as to promote

[9] See Politics, Religion, and Culture | TikTok Advertising Policies. See also TikTok's Community Guidelines on Civic and Election integrity.

[10] See https://ads.tiktok.com/help/article/tiktok-ads-policy-politics-religion-and-culture.

[11] See https://www.tiktok.com/legal/page/global/bc-policy/en.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013648

public health initiatives like COVID-19 vaccination programs). GPPPAs are also ineligible for Series,[12] the Creator Rewards Programme[13] and receiving Gifts[14] on their content.

42. Under the Political Ads Policy:

    i. Under the heading "***Political content***", TikTok explains that it prohibits: (i) all Ads that feature political content; (ii) all accounts that are identified as GPPPA accounts from posting Ads (except in limited circumstances, e.g. to promote public health initiatives like COVID-19 booster programs, in which case the accounts are required to work with a TikTok representative); and (iii) all advertisers from referencing, promoting, or opposing candidates, legislation, policies, etc. For instance:

    > "*We do not allow ads featuring political content across any of our monetization features, including paid ads, creators being paid to make branded political content, and other promotional tools on the platform. We also impose prohibitions at the account level for advertisers we identify as politicians and political parties, who are not able to create advertising accounts or access advertising features.*"

    The policy links to the GPPPA Policy Help Centre Article[15] which outlines what those account level restrictions are.

    ii. The Political Ads Policy then goes on to explain that the types of advertisers that are not allowed to place Ads include:

    > "*candidates or nominees for public office, political parties, elected or appointed government officials;*
    >
    > *the spouses of candidates, elected, or appointed government officials with official duties or offices; and*
    >
    > *royal Family members with official government capacities*"

    iii. Under the heading, "*Examples of **what** is not allowed*", the policy lists:

    > "*referencing, promoting, or opposing candidates or nominees for public office, political parties, or elected or appointed government officials;*
    >
    > *advocacy for or against past, current, or proposed referenda, ballot measures, and legislative, judicial, or regulatory outcomes or processes. This includes ads that promote or attack government policies or track records;*
    >
    > *referencing, promoting, or selling merchandise that features prohibited individuals, entities, or content, including campaign slogans, symbols, or logos; and*
    >
    > *references to an election, including voter registration, voter turnout, and appeals for votes, such as ads that encourage people to vote.*"

43. Under the Branded Content Policy:

    i. Under the heading "*What is Branded Content?*", TikTok explains that:

---

[12] See https://support.tiktok.com/en/business-and-creator/tiktok-series/about-tiktok-series.
[13] See https://support.tiktok.com/en/business-and-creator/creator-rewards-program/creator-rewards-program.
[14] See https://support.tiktok.com/en/business-and-creator/video-gifts-on-tiktok/video-gifts-on-tiktok.
[15] See https://support.tiktok.com/en/using-tiktok/growing-your-audience/government-politician-and-political-party-accounts.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

"[Branded Content] *is* ***content that promotes a third party brand or its products or services in exchange for payment or any other incentive****. Branded Content includes content which features or references:*

- *A product or service that has been gifted to you by (or on behalf of) a brand;*
- *A brand, product or service that you have been paid to post about (whether in the form of money or a gift);*
- *A product or service for which you will receive a commission on any sales (for instance, via an affiliate link or promotional code); or*
- *A brand that you have or have had a commercial relationship with, such as acting as a brand ambassador.*"

ii. Under the headings "*What do you need to do when you post Branded Content?*" and "*Prohibited Industries*", TikTok explains that:

"***Political content - All political or issue-based advertising is prohibited****:*

- *Candidates or nominees for public office, political parties, elected or appointed government officials are prohibited from advertising.*
- *The spouses of candidates, elected, or appointed government officials with official duties or offices are prohibited from advertising.*
- *Royal Family members with official government capacities are also prohibited from advertising.*
- *Political and issue-based advertising includes:*
  - *ads that reference, promote, or oppose candidates or nominees for public office, political parties, or elected or appointed government officials;*
  - *ads that reference an election, including voter registration, voter turnout, and appeals for votes;*
  - *ads that include advocacy for or against past, current, or proposed referenda, ballot measures, and legislative, judicial or regulatory outcomes or processes, including ads that promote or attack government policies or track records; and*
  - *ads that reference, promote, sell, merchandise that feature prohibited individuals, entities or content, including campaign slogans, symbols, or logos.*"

"*Other prohibited products or services - Any other businesses, products, or services that are highly controversial, distasteful, violent, or dangerous would be deemed incompatible with the TikTok platform. Examples include… Branded Content that refers to a sensitive event or topic (e.g. deaths, natural/industrial disasters, violent attacks, civil disorder, etc.) and contain any of the following:…* ***political campaigning****…*"

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013650

44. See further TikTok's responses to ***questions 5(a)***, ***(b)***, and ***(f)*** of the Recommender Systems & Elections RFI and GPPPAs Help Centre article.[16]

b. Please provide a detailed description of the measures and internal controls in place to detect "political content" in advertisements and enforce the provider of TikTok's prohibition on political advertising.

45. At the outset, TikTok respectfully disagrees with the suggestion in the RFI that "*Based on publicly available information, it appears that the provider of TikTok may not have effectively implemented appropriate content moderation processes to enforce its terms and conditions, in particular in relation to the restrictions applicable to 'political advertising'*" (RFI, page 3). As the Commission is aware, Article 8 and Recital 30 DSA expressly state that the DSA does not impose, and nothing in it should be construed as imposing, a general monitoring obligation or a general active fact-finding obligation on providers, nor a general obligation for providers to take proactive measures in relation to illegal content. The DSA likewise does not include an obligation for providers to generally monitor for policy-violating content.

46. Notwithstanding the foregoing, TikTok takes measures to detect/prevent political content in Ads at both the account level and content level, both before publication and (for Ads that are initially approved) after.

    i. First, as explained in response to ***question 8(a)*** above, accounts which TikTok identifies as GPPPAs are, with very limited exceptions, prevented from accessing TikTok's monetisation features (including Ads). The way TikTok identifies GPPPAs is explained in response to ***question 5(a)*** of TikTok's response to the Recommender Systems & Elections RFI.

    ii. Second, as also explained in response to ***question 8(a)*** above, no advertisers are allowed to, reference, promote or oppose candidates, past or future legislation, government policies or track records, refer to elections or election registrations, or promote merchandise that features prohibited individuals, entities or content, such as campaign symbols, symbols or logos. TikTok identifies, and enforces its prohibition against, any such Ads content via its pre-publication and post-publication detection and enforcement mechanisms. Specifically:

        - Before Ads are published on TikTok, they undergo a review process to ensure compliance with applicable policies, including the Political Ads Policy. TikTok's automated detection systems look for any violations of TikTok's Political Ads Policy based on a wide range of indicators. For example, these automated detection systems assess whether Ads contain certain political elements or indicators – based on keyword detection (e.g., checking titles, captions and in-video text) and image and video detection (e.g., checking images in a video). If the systems conclude that the Ad violates the Political Ads Policy with a sufficiently high degree of certainty, it will be rejected (i.e., not published on the Platform); if the systems have a lesser degree of certainty, the Ad will be sent for review by a human moderator.

        - After an Ad has been published, if it reaches a certain high number of views, it is subject to further review by TikTok's automated mechanisms, to detect any violations of TikTok's Ads Policies (including TikTok's Political Ads Policy). This may trigger further review by a human moderator.

[16] See https://support.tiktok.com/en/using-tiktok/growing-your-audience/government-politician-and-political-party-accounts.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013651

c. Please provide a detailed description of the measures and internal controls in place to enforce TikTok's policy prohibiting "*creators from being paid to make branded political content*". Please provide all internal documents relating to the controls in place to enforce this policy. In addition, please explain whether the provider of TikTok has put in place any specific measures to detect branded political content being livestreamed ([17]).

47. TikTok understands that, in this ***question 8(c)***, the Commission is asking whether TikTok has put in place measures to detect political content in undeclared Branded Content (including in LIVE content) in order to enforce its Branded Content Policy. TikTok first notes that, e.g., Articles 8 and 26 DSA make clear that providers' obligation is to provide users with a functionality to declare commercial communications – not to ensure that users do so, or to proactively seek out commercial communications that have not been declared by users.

48. Branded Content is distinct from Ads, in that Ads are inherently declared as such (because TikTok directly receives payment for them). TikTok's measures for detection of political content in Ads are described in response to ***question 8(b)*** above. While TikTok's Trust and Safety team makes efforts to identify videos containing Branded Content that have not been declared as such by the poster, this can be challenging in practice.[18] In many cases, it is difficult for TikTok to discern or discover an undeclared off-Platform payment or incentive that a user receives directly from a third party for posting videos (or streaming LIVE content) containing undeclared Branded Content. The challenge is even greater with LIVE content, which is time-sensitive. As explained above, TikTok allows political content in LIVE content that is not Branded Content.

49. Accordingly, TikTok provides, and TikTok's Terms of Service and Branded Content Policy require users to use, the commercial content disclosure toggle[19] so that, whenever they post Branded Content, users can declare it to TikTok. Branded Content must comply with the Branded Content Policy (in addition to the Community Guidelines) which, as explained in response to ***question 8(a)*** above, prohibits political content.

50. If the disclosure toggle has been used to declare a posted short video as Branded Content, TikTok moderates it specifically against the Branded Content Policy using a combination of machine and human moderation strategies. TikTok uses the same political content machine detection model for Branded Content short video posts as for Ads.

51. If a user violates TikTok's Branded Content policy and does not disclose to TikTok that their posted video is Branded Content, the video is still moderated against the Community Guidelines in accordance with TikTok's content moderation policies that apply to all uploaded content. If TikTok subsequently identifies that such content is undisclosed Branded Content (i.e., where the user has received payment or another incentive but has not disclosed this to TikTok), TikTok will then re-moderate that content as Branded Content, under the Branded Content Policy.

[17] The Commission notes that there are media reports mentioning that the influencer BOGPR might have paid recipients of the service hosting livestreams in exchange for their political supports of one candidate to the Romanian Presidential election (see Le Monde, "*En Roumanie, derrière le succès politique de Calin Georgescu, l'ombre d'un mysterieux influenceur*", 28 November 2024).

[18] In certain cases, such as through TikTok's CIO investigations, TikTok may identify campaigns or other coordinated activities which involve users posting political content for payment or other incentives.

[19] See https://support.tiktok.com/en/business-and-creator/creator-and-business-accounts/promoting-a-brand-product-or-service.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

d. For content moderation and any algorithmic systems employed to detect "political content" and enforce the above-mentioned policies, please provide a detailed description and performance metrics about how the provider of TikTok monitors the performance of the underlying models in production, including data drift tracking, models' accuracy, precision, recall at the time of training, and their accuracy, precision, recall from 1 September 2024 to 24 November 2024, including for the detection of political content in Romanian.

Please include any documents presenting the results of such monitoring from 1 September 2024 to 24 November 2024.

52. With respect to TikTok's machine learning models which detect political content in promote and auction ads[20] (the "**political content models**"), TikTok notes that:

i. For TikTok's political content models in Romania from 1 September to 24 November 2024, the political content models':

- recall rate (i.e., percentage of Ads identified by automated moderation as containing suspected political content that were then recalled for human moderator review, in both pre-posting and post-posting stages) was 100%; and
- precision (i.e., out of all recalled Ads, the percentage of Ads confirmed as containing political content violations) was 78.1%.

In terms of detection accuracy, the metric for this mitigation measure is proactive detection rate (i.e. the proportion of violating Ads that TikTok detects and removes before they receive any views), which was 99.5%.

ii. With respect to how TikTok monitors the accuracy or performance of the political content models, TikTok monitors on a daily basis both the hit volume (i.e. the total number of times an Ad is served or displayed during a specific period), and rejection rate (i.e., how many Ads are rejected from the Platform on the basis that they amount to political ads), and then optimises the model based on the rejection rate. TikTok also reviews e.g., samples of posted content for quality control purposes, and user reports in respect of political content which violates TikTok's policies, to ensure the coverage and accuracy of the political content models' detection capability.

iii. With respect to how the political content models "learn", or have their datasets updated, TikTok has automatic daily processes which:

- identify information (such as certain text and images) in content that has been identified as violative of TikTok's Political Ads Policy; and
- collect publicly available information about public affairs and political events (which is subject to a compliance review),

and that is used to update the political content models to ensure they are kept up-to-date.

iv. With respect to metrics at the time when the political content models were trained. TikTok updates the political content models on a regular basis. The last major update was in Q3 2024, the average proactive detection rate globally was 99.37% (98.5% in Romania), with approximately 100% recall rate and approximately 42% precision rate.

v. As part of TikTok's work to continuously improve its suite of mitigation measures (including to take into account the evolution of new and existing risks), TikTok is developing strategies to enhance its political content models' ability to detect potentially violative content at an early stage in less widely-spoken languages, including Romanian. Such initiatives include increasing

[20] In the time available, TikTok has not been able to obtain metrics in respect of its detection of political content in Branded Content.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013653

the availability of machine-readable transcripts of the visible text within the video (created using optical character recognition (OCR)) and machine-readable transcripts of audio or music within a video (created using automatic speech recognition (ASR)). In the meantime, TikTok is also lowering the thresholds at which political content is considered potentially violative by the political content models, so that more content is sent for human review.

9. The provider of TikTok's Community Guidelines does not allow "covert influence operations" which are defined as "*coordinated, inauthentic behaviour where networks of accounts strategically work together to mislead people or our systems and influence public discussion. This can include attempting to undermine the results of an election, influencing parts of an armed conflict, or shaping public discussion of social issues.*" ([21])

a. Please provide details and internal documents on the detection and actions taken against covert influence operations in the context of the Romanian elections and the results of those actions taken, including whether the provider of TikTok identified signals of off-platform deceptive coordination to detect any covert influence operation in the context of Romanian elections.

53. TikTok understands that, in this ***question 9(a)***, the Commission is asking about CIOs that TikTok identified in the context of the Romanian elections.

54. TikTok has provided details regarding its on- and off-Platform detection of, and actions taken against, CIOs in the context of the Romanian elections, and the results of those actions, in its responses to ***question 1*** of ***Section 1.1*** and ***question 6*** of ***Section 1.2*** of the Second Romanian Elections RFI. Please also see TikTok's responses to ***questions 5(c)*** and ***11(b)*** of the Recommender Systems & Elections RFI, in which TikTok provides details regarding its activities to detect and disrupt CIOs in general.

55. As the Commission is aware, TikTok has also provided details of how it counters a broad range of inauthentic and deceptive behaviour on the Platform using a variety of measures: (i) in its communications with the Commission; (ii) in response to the Recommender Systems & Elections RFI (in particular ***question 5***); and (iii) in public articles published on TikTok's website, e.g.,:

    i. https://newsroom.tiktok.com/en-eu/how-tiktok-counters-deceptive-behaviour;

    ii. the Newsroom Post;[22] and

    iii. TikTok's actions against inauthentic activities in the context of elections include: (i) defining CIOs, (ii) building an influence operations taskforce; (iii) disrupting and removing networks which violate TikTok's covert influence policies; (iv) targeting inauthentic expression; (v) countering fake engagement; and (vi) safeguarding authentic experience. In addition to this wide range of proactive detection measures, as the Commission knows, TikTok includes mechanisms which enable its users to easily report content or accounts that they are concerned about.

56. In the context of the Romanian elections, as just one of its measures to ensure election integrity, TikTok has identified and disrupted (e.g., by banning accounts involved and removing their content) the following CIOs:

    i. **A network of 22 accounts, with 299,692 followers,** operating from Romania that attempted to target Romanian audiences in order to spread misinformation and promote narratives critical of the current Romanian government and in support of AUR party, which TikTok transparently reported in September 2024.[23]

    ii. **A network of 78 accounts operating from Romania, with 1,781 followers**, that attempted to promote the candidate Călin Georgescu on TikTok. TikTok assessed that this CIO network was

[21] See Community Guidelines | TikTok.
[22] See https://newsroom.tiktok.com/en-eu/continuing-to-protect-the-integrity-of-tiktok-during-romanian-elections.
[23] See https://www.tiktok.com/transparency/en/covert-influence-operations.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

operating from Romania and disrupted it in late November 2024. This will be reflected in TikTok's November Covert Influence Operations Report.

iii. **A network of 12 accounts operating from Romania, with 1,704 followers**, that attempted to amplify narratives in Romanian favourable to independent political candidate Mircea Geoană. TikTok assessed that this network was operating from Romania and disrupted it in late November. This will also be reflected in TikTok's November Covert Influence Operations Report.

iv. **TikTok notes that it also operates broader defences against covert influence campaigns across Europe.** In September 2024, TikTok took global action against a covert network linked to Russian state-owned news agency Sputnik Media. When TikTok removes such networks, it continues monitoring for any attempts to re-establish them. As part of TikTok's anti-recidivism strategy, it removed 11 accounts in November 2024 that TikTok believed to be associated with Sputnik Media and targeting Romanian and Moldovan audiences.

57. In addition to undertaking investigations into potential CIOs identified through its own internal detection efforts, TikTok also undertakes investigations into potential CIOs following external reports, such as reports received from relevant national authorities.

58. For example, on 31 October 2024, TikTok received a request from the Romanian National Police and ANCOM (please see TikTok's response to ***question 3*** above) to investigate certain inauthentic accounts and botnets. In response to that request, TikTok's intelligence ("**IOID**") and investigations ("**IOTD**") teams conducted further investigations and, in late November 2024, successfully disrupted a CIO network of 12 accounts operating from Romania, with 1,704 followers, that attempted to amplify narratives in Romania favourable to the independent political candidate Mircea Geoană. This will also be reflected in TikTok's November Covert Influence Operations Report.

59. TikTok intends to provide the Commission with further information in its Response to the Second Romanian Elections RFI regarding instances that it investigated in the context of the Romanian elections which did not constitute CIOs.

b. Please explain all subsequent analysis conducted by the provider of TikTok on the occurrence of covert influence operations in the context of the Romanian elections so far. Please provide all internal documents showing how it was decided to pursue the analysis, how and when the analysis was conducted as well as the results of such analysis.

60. Please see TikTok's response to ***question 9(a)*** above, as well as TikTok's response to ***question 1*** of Ref. Ares(2024)8716531.

c. Please describe whether the provider of TikTok has put in place any mitigation measures against authentic accounts using deceptive tactics in a coordinated way to influence public discussions during elections, such as astroturfing. If yes, please describe these measures and provide all internal documents reflecting the results of the application of such mitigation measures during the Romanian elections. In addition, please explain how the provider of TikTok distinguishes between authentic accounts using deceptive tactics in a coordinated way to influence public discussions during elections, and legitimate coordinated political campaigns.

61. If this question is referring to authentic accounts that post organic "*political content*" but where the user receives payment or other incentive in return for posting such content (i.e. content covered by the Branded Content Policy), this constitutes a violation of such policy and is discussed in TikTok's response to ***question 8*** above. Alternatively, if this question is referring to TikTok's mitigation measures specifically in respect of CIOs, TikTok responds as follows.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013655

62. TikTok's Transparency Centre explains[24] how TikTok defines, assesses and counters CIOs. In addition to those CIO policies, TikTok's Integrity & Authenticity Community Guidelines[25] prohibit deceptive behaviour and organised attempts to mislead or deceive TikTok's users. The Community Guidelines also explain that CIOs can include coordinating content or engagement activity to artificially promote certain narratives or trends.

63. To enforce these policies, TikTok has a team of dedicated internal specialists who focus specifically on detecting and disrupting CIOs. TikTok is aware that individual content-level mitigation measures, such as content moderation, can be insufficient to combat CIO behaviours and coordination. However, while the harmful effects of multiple accounts engaging in coordinated deceptive behaviour can be significant, the potential impact on users' freedom of expression if enforcement action (such as account bans) is wrongly taken against multiple (authentic) accounts can also be significant – and distinguishing between such (authentic) accounts and behaviours is not straightforward. TikTok's CIO investigations are resource-intensive and require in-depth analysis. As explained further below, such investigations focus on behaviour and assessing linkages between accounts and techniques to determine if actors are engaging in a coordinated effort to mislead TikTok's systems or its community. In this regard, TikTok uses several types of information (including open-source and proprietary) to assess CIOs, and leverages a standard framework of confidence assessment to help ensure that it makes consistent and accurate determinations. Where TikTok's teams have the necessary high degree of confidence that an account is engaged in or connected to networks previously removed as part of a CIO, it is removed from the Platform.

64. For conduct to amount to a CIO and be actionable under TikTok's CIO policy, TikTok must be able to establish evidence of three particular elements. The combination of these elements allows TikTok to conclude that there is a CIO, whereas the absence of one or more elements might indicate a legitimate coordinated political campaign:

    i. Coordination between multiple accounts: TikTok's specialist CIO investigators carefully look for and assess technical and other indicators of coordination (e.g. shared devices or IP addresses, or similar user handles and/or profile pictures across accounts). It is the 'coordination' element that means that activity falls within TikTok's CIO policy and not (only) other policies within the Community Guidelines that are limited to individual accounts – such as those prohibiting impersonation (see further TikTok's response to ***question 5(c)*** of the Recommender Systems & Elections RFI).

    ii. Misleading/deceiving TikTok's systems or users: For this element, TikTok's specialist CIO investigators look for technical and other indicators of attempts to mislead, or to conceal behaviour or identities (e.g. use of proxy IPs and VPNs to attempt to post content in another market, or impersonation including use of fake personas). Behaviour may also be considered misleading where its true objective is concealed. It is the 'misleading' element that means that activity falls within TikTok's CIO policy and is not considered genuine political discourse or social activism (which may involve genuine and non-deceptive coordination to influence public debate). TikTok can also assess for astroturfing behaviour as part of its standard procedures in investigations into CIOs, should it be relevant to a given investigation.

    iii. Attempt to manipulate (or corrupt) public debate: For this element, TikTok's specialist CIO investigators assess for a cohesive strategic aim behind an apparent network of accounts, to identify what (if anything) the network is attempting to influence. TikTok then weighs up the evidence to form a holistic assessment. It is the 'strategic intent to manipulate' element that separates CIOs from other forms of inauthentic behaviour that are (separately) prohibited under the Community Guidelines (e.g., spam, which might be coordinated and misleading but is not designed with the strategic aim to manipulate public debate).

65. As the Commission has understood, the covert influence operations (CIO) efforts are one of the ways to counter deceptive behaviour on the platform. There is a wide-range of deceptive behaviour carried out online across all platforms, which can touch on important social issues including elections—from spam

[24] See https://www.tiktok.com/transparency/en-us/countering-influence-operations.
[25] See https://www.tiktok.com/community-guidelines/en/integrity-authenticity.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

to impersonation to covert influence operations. In Romania specifically, TikTok proactively prevented nearly 45 million fake likes, more than 27 million fake follow requests, and blocked more than 400 thousand spam accounts from creation since September 2024 alone. TikTok proactively takes action against these inauthentic activities, which violate its policies. TikTik has summarised its efforts in countering deceptive behaviour more broadly in this recent post: https://newsroom.tiktok.com/en-eu/how-tiktok-counters-deceptive-behaviour.

d. Based on public reports ([26]), a number of accounts, so-called 'murky' accounts, have been detected on TikTok and flagged to the service. Please explain what concrete steps TikTok has taken vis-à-vis these accounts.

Please provide all internal documentation and exchanges referring to these accounts.

66. Please see TikTok's responses to ***questions 9(a)-(c)*** above, which explain TikTok's approach to CIOs.

67. With respect to the specific accounts reported to TikTok by Democracy Reporting International ("**DRI**") in relation to the Romanian elections (including in the document at footnote 15 of the RFI), TikTok received two reports from DRI reporting specific accounts. Both of these reports were carefully assessed by TikTok's specialist teams, including TikTok's Incident Management (IM) team.

    i. The first DRI report[27] was received on 28 November 2024. DRI reported 66 links (including one duplicate link), referring to 65 TikTok accounts. Of those:

        - 61 were permanently banned for violating TikTok's Community Guidelines prohibiting impersonation. TikTok had already identified and taken enforcement action with respect to a number of these, and it investigated and took action against the others following receipt of DRI's report.

        - The remaining 4 accounts were clearly disclosed as 'fan' accounts, and were therefore assessed as not violating the Community Guidelines. Please see further TikTok's response to ***question 5(c)*** of the Recommender Systems & Elections RFI.

    ii. The second DRI report[28] was received on 4 December 2024. DRI reported 48 links, referring to 48 TikTok accounts. Of those:

        - 42 were permanently banned for violating TikTok's Community Guidelines prohibiting impersonation. TikTok had already identified and taken enforcement action with respect to a number of those, and it investigated and took action against the others following receipt of DRI's report.

        - The remaining 6 accounts were clearly disclosed as 'fan' accounts, and were therefore assessed as not violating the Community Guidelines. Please see further TikTok's response to ***question 5(c)*** of the Recommender Systems & Elections RFI.

68. TikTok's IOTD team, specialist CIO investigators also performed an in-depth analysis and found no evidence of any potential CIO networks with respect to these specific accounts reported by DRI. If the Commission is referring to other 'murky accounts' please provide details or the source.

---

[26] See https://democracy-reporting.org/en/office/EU/publications/manufactured-support-how-inauthentic-activity-on-tiktok-bolstered-the-far-right-in-romania.

[27] See footnote 26 above.

[28] See https://democracy-reporting.org/en/office/global/publications/skirting-the-rules-the-romanian-far-right-continues-to-enjoy-inauthentic-and-prohibited-support-on-tiktok.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013657

10. In its Guidelines on the mitigation of systemic risks for electoral processes, the Commission recommends that providers of VLOPs and VLOSEs, in addition to their legal obligations under Article 40 of Regulation (EU) 2022/2065, "*provide free access to data to study risks related to electoral processes, including, where necessary, those not available on the VLOPs and VLOSEs interface, to relevant third-party stakeholders*." Please clarify whether the provider of TikTok has put in place additional and tailor-made tools, or features, to enable a wider range of third parties, in addition to qualified researchers under Article 40(12) of Regulation (EU) 2022/2065, to have access to publicly accessible data to detect, identify and understand systemic risks related to electoral processes. If yes, please provide details of such features.

69. TikTok understands that this ***question 10*** refers to paragraph 30 of the Elections Guidelines, which discusses "*additional and tailor-made tools, or features*" in addition to those that are required for a provider to comply with Article 40(12), and states that "[a]*ccess to such tools or features could be extended to a wider range of third parties*".

70. TikTok has not put in place such "*additional and tailor-made*" tools or features. However, TikTok has previously provided the Commission with detailed information concerning its Research Tools and Virtual Compute Environment. Please see further:

    i. TikTok's responses to RFI Ref.Ares(2024)398592 (the "**Public Data Access RFI**");

    ii. TikTok's responses to questions 68-73 of RFI Ref.Ares(2024)2393266 (the "**Youth Investigation RFI**"); and

    iii. TikTok's resources for researchers at https://developers.tiktok.com/products/research-api/.

71. In addition, TikTok's Codebook[29] guides researchers as to the data available and the functionality that TikTok provides to researchers through its Research Tools functionality.

72. TikTok notes that tools that it already makes available to researchers are being used to study electoral processes and civic engagement, including in 41 projects across the EU alone (with more across the rest of Europe). This includes three ongoing research projects based in Romania, which focus on:

    i. misinformation and disinformation in Romania – with a focus on Romanian misinformation trends, in particular with respect to influencers who post content relating to socio-political issues;

    ii. participation in democracy and protests in Romania and other European countries, including taking into account how social actors leverage personal networks on social media to attain movement outcomes such as participant mobilisation and organisation, collective action or commitment to activism; and

    iii. the impact of negative portrayals of minorities in Romanian media and social media.

73. TikTok continues to constructively engage with researchers in accordance with its Research Tools policy and terms[30] and, to this end, is currently processing a research access application seeking to understand and study content creators and their engagement during the 2024 Romanian elections.

74. TikTok also notes that it makes public aggregated data available through its Virtual Compute Environment to certain non-academic, not-for-profit organisations which are not "researchers" within Article 40 DSA. As noted above, please see TikTok's responses to RFI Ref.Ares(2024)398592 (the "**Public Data Access RFI**") and TikTok's responses to questions 68-73 of RFI Ref.Ares(2024)2393266 (the "**Youth Investigation RFI**") for more detailed information.

[29] See https://developers.tiktok.com/doc/research-api-codebook?enter_method=left_navigation.
[30] See https://developers.tiktok.com/products/research-api/.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013658

11. In its Guidelines on the mitigation of systemic risks for electoral processes, the Commission recommends that "*it is of particular importance that providers of VLOPs and VLOSEs show that content moderation decisions do not affect the equality of candidates or disproportionately favour or promote voices representing certain (polarised) views*". Please clarify whether the provider of TikTok has any measures in place to ensure that content moderation decisions do not affect the equality of candidates or disproportionately favour or promote voices representing certain (polarised) views. If yes, please provide a detailed explanation of these measures.

75. TikTok understands that this ***question 11*** is asking whether TikTok has implemented content moderation-related processes and controls to ensure that all candidates are treated in an equal manner, so as not to "*affect the equality of candidates*", and to ensure that rights to freedom of expression are respected so as not to disproportionately favour or promote voices representing certain (polarised) views. In this regard, please see further TikTok's response to ***question 2*** of the Recommender Systems & Elections RFI.

76. As explained in the Recommender Systems & Elections RFI, TikTok recognises the risk of potential bias in its content moderation systems. Committing to impartiality, evidence-based decision making and fair outcomes represent one of TikTok's core eight guiding community principles.

77. Decisions based on the Community Guidelines aim to achieve equal treatment of content and prohibit preferential treatment of any candidate, party or political group. This is clearly stated in the Community Guidelines:

> "*TikTok has eight guiding community principles that are grounded in safety and our commitment to respecting human rights. Our principles shape our day-to-day work and guide how we approach difficult enforcement decisions. They are centered on these themes:*
>
> - *Balancing harm prevention and expression*
> - *Embracing human dignity*
> - *Ensuring our actions are fair*
>
> [...]
>
> *2.* ***Enable free expression:*** *The creativity unlocked by expression is what powers our vibrant community. We honor this principle by providing the opportunity to share freely on our platform while also proactively addressing behavior that can inhibit speech of others. However, free expression is not an absolute right – it is always considered in proportion to its potential harm, and does not extend to having your content recommended in the For You feed.*
>
> [...]
>
> *8.* ***Be fair and just:*** *Moderating millions of pieces of content each day is a complex effort, and developing a trusted process to do so is foundational. We are committed to being impartial and evidence-based, producing fair outcomes, giving notice of enforcement actions, and providing an opportunity to appeal.*"

78. At times, these principles may be in tension with each other, and TikTok carefully considers this when weighing one principle over another. These considerations are informed by international legal frameworks and industry best practices, including the UN Guiding Principles on Business and Human Rights,[31] the International Bill of Human Rights,[32] the Convention on the Rights of Children,[33] and the Santa Clara Principles on Transparency and Accountability in Content Moderation. TikTok also seeks input from its community, safety and public health experts, and Advisory Councils.[34]

---

[31] See https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf.
[32] See https://www.ohchr.org/en/what-are-human-rights/international-bill-human-rights.
[33] See https://www.ohchr.org/en/professionalinterest/pages/crc.aspx.
[34] See https://www.tiktok.com/transparency/advisory-councils/?lang=en.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013659

79. In implementing measures that are reasonable, proportionate and effective, and that do not unduly impact on freedom of expression, TikTok employs the following measures.

Policy development and appropriate exemptions

80. As noted above, TikTok's Community Guidelines are the starting point when it comes to how TikTok enforces its content moderation strategies and practices. While preserving the ability for users to express themselves, one of TikTok's highest priorities is to prevent harmful misinformation, including election misinformation.

81. TikTok's approach to content moderation applies the same criteria, regardless of who creates the content being moderated. However, in some situations, it may be in the public interest to be able to view content that would otherwise violate TikTok's policies. These are described in as "*Public Interest Exemptions*" in the Enforcement section[35] of the Community Guidelines:

> "*Public interest refers to topics that inform, inspire, or educate the community and enhance deliberation about matters of broad collective significance. We may allow content to remain on TikTok under one of these public interest exceptions:*
>
> - *Documentary*
> - *Educational*
> - *Medical and Scientific*
> - *Counterspeech*
> - *Satirical*
> - *Artistic*"

82. In the specific context of Election Misinformation, examples of exceptions that are permitted on the grounds of freedom of expression under the Community Guidelines may include, for example:

    i. statements of personal opinion (provided it does not include harmful misinformation) – e.g., "*I don't like this candidate*" or "*I cannot support this party*", provided the claim does not include verifiably false information or content that misleads users about the civic process in question; or

    ii. content that is ineligible for the 'For You feed' if it contains unverified claims about the outcome of an election.

83. TikTok takes a holistic approach when updating the Community Guidelines, by engaging in consultations with both internal and external subject matter experts, including through TikTok's Safety Advisory Councils. For example, as explained in TikTok's 2024 Risk Assessment Report, TikTok updated the Community Guidelines in May 2024 following individual consultations with over 40 global experts in democracy, tech policy and freedom of speech. The update added new policies regarding unverified election claims, election misconduct, distribution of hacked materials, and misrepresented civic sources. In addition, TikTok's Platform Fairness team was consulted on the potential human rights impact of the update.

Localised approach

84. Through TikTok's Regional Policy teams, OPM engagements and other outreach initiatives, TikTok takes steps to understand local and regional context, complexities and nuances, and the experiences of particular communities, and collaborates with partners across industry sectors, before making decisions that disproportionately affect those communities. These initiatives provide insight, cultural context, and local knowledge. They also inform TikTok's understanding of the interplay between global and local policies,

[35] See https://www.tiktok.com/community-guidelines/en/enforcement?cgversion=2023#:~:text=Notice%20and%20Appeals-,Enforcement,-Released%20April%2017.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013660

and help to drive local discussions and engagement on content policy issues, ensuring that TikTok does not take measures that may have disproportionate impacts on freedom of expression.

85. In the context of the Romanian elections specifically, TikTok engaged extensively with local authorities, fact checkers and civil society organisations to ensure preparedness for these elections. Please see further TikTok's responses to ***questions 5*** and ***6*** above, and TikTok's Newsroom Post, for more details.

Expert teams with specialist training and resources

86. As explained in response to ***question 7*** above, TikTok has also developed Moderator Guidelines in relation to elections content for EU countries (including Romania) for the purposes of processing illegal content reports under the DSA. These include topics such as election silence periods, the sharing of opinion polls, ballot papers and support for candidates.

Elections Integrity Programme

87. Moreover, TikTok has developed an Elections Integrity Programme which is designed to develop and utilise expertise in tackling Election Misinformation on a global basis, and which can be applied at a regional/local level. The Romanian elections are included in this Elections Integrity Programme. The Programme enables TikTok's regional/local team to draw on and benefit from experience, expertise and insights gained during elections in other regions, as part of TikTok's commitment to continuous improvement. The Programme includes the specific measures listed below.

88. **Specialist Integrity Moderators**: As noted above, TikTok has established dedicated Integrity Moderator teams which are made up of experienced moderators to moderate suspected misinformation content. These specialists undergo specialised training on Integrity & Authenticity policy issues, with a specific focus on how to differentiate content with freedom of speech value vs actual misinformation, which contributes to ensuring expert decision-making on misinformation issues. The teams are also subject to ongoing quality assurance, which ensures that accurate decisions are taken and feedback is provided.

89. **Platform Fairness**: The Platform Fairness team is a horizontal-issue policy team committed to equitable treatment across product policies, moderation processes, models and in-app features. Their approach is grounded in fundamental universal human rights (UDHR Article 1, 2) and ensuring that the Platform supports/protects vulnerable and marginalised groups, taking an equity-based approach informed by targeted universalism.

90. With respect to risks to elections and civic integrity, the Platform Fairness team has:

    i. consulted on the launch of the new election policies to ensure fairness in policy language (such as the update of the Community Guidelines in May 2024 that added new policies and enforcement regarding unverified elections claims, election misconduct, distribution of hacked materials, and misrepresented civic sources);

    ii. designed and deployed implicit bias training for all moderators to mitigate against bias in scaled moderation enforcement outcomes. The Moderation Fairness (Implicit Bias) Training Course is designed to strengthen the moderator mindset to improve quick decision-making in cases where bias might be a factor; and

    iii. revamped the moderation fairness course materials to include a Level 3 course with a specific module addressing moderating fairly on political and religious content. This course is scheduled to be deployed in EMEA in January 2025.

91. **Fact-checking programme**: TikTok works with International Fact-Checking Network (IFCN) certified fact-checkers as part of its Fact-Checking Programme to ensure that there is a robust process before borderline content (see further TikTok's response to ***question 6*** of the Recommender Systems & Elections RFI) is moderated on the Platform, striking a balance between tackling Election Misinformation and protecting freedom of expression. This is supported by TikTok's maintained repository of fact-checked information (such as claims, statements and conspiracy theories that have been fact-checked by TikTok's

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

external fact-checking partners). For further detail on TikTok's fact-checking programme in the context of Romania, please see TikTok's response to ***question 4*** above and TikTok's Newsroom Post.

92. **Review of GPPPAs content**: With respect to the content created by GPPPAs, TikTok reviews such content to avoid mis-moderation issues (i.e. to avoid bona fide content being removed, with consequent negative impact on freedom of speech, which could be significant in an election context). While TikTok uses the same content moderation practices for GPPPA content as for other TikTok accounts, additional checks are also implemented before such accounts are temporarily banned or content removed.

Requests from government authorities

93. Finally, as explained in response to ***question 3*** above, in addition to its standard reporting mechanisms, TikTok also has exchanged contact information directly with relevant government authorities (including in Romania) to provide dedicated channels for them to report suspected illegal content, including for violations of election legislation.

94. Once TikTok receives such requests, they are managed under TikTok's Global Policy Framework on Content Removal Requests. This ensures that such requests (and related escalations) are handled in a balanced, compliant, and principled manner – aligning with the Community Guidelines, TikTok's other policies, applicable law, and TikTok's commitments to human rights principles.

12. The provider of TikTok reported in the risk assessment 2023 ([36]) that mitigation measures in relation to elections "*may involve a targeted sweep focused on use of a specific keyword, URL or hashtag identified as being associated with high-risk content, such as where certain content is "viral*". It is also reported that such content may then undergo Manual Review in order to detect and mitigate safety risks, including in relation to Election Misinformation." In the context of the Romanian elections, please provide:

a. A list of keywords, URLs and hashtags identified as being associated with high-risk content from the 1 September to 24 November 2024;

95. TikTok adds keywords and hashtags that it identifies as 'high risk' (i.e., associated and/or connected to content that was found to be violative) to its wider keyword and hashtag list on an iterative basis, in light of dynamic external changes and trends identified on the Platform. 'High risk'-related keywords and hashtags identified in the context of the Romanian elections during the Relevant Period were as follows:

Keywords

i. Kill – Ucide / Omor
ii. To Kill – A ucide / A omorî
iii. Civil War – Război civil
iv. War on Lasconi – Război împotriva lui Lasconi / Război contra lui Lasconi
v. War on Georgescu – Război împotriva lui Georgescu / Război contra lui Georgescu
vi. Target Politicians – Ţinteşte politicieni / Vizaţi politicieni
vii. Kill Politicians – Ucide politicieni / Omorâţi politicieni
viii. Război hibrid (hybrid war)
ix. Război mondial (World war)
x. Declaraţie de război (war declaration)
xi. România război (Romania war)
xii. România în război (Romania at war)

Hashtags

xiii. #razboi
xiv. #război
xv. #razboiromania

[36] Risk assessment 2: integrity & misinformation. Module 1: risks to election and civic integrity, p. 22.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

xvi. #rascoala
xvii. #răscoală

96. Following the extremely volatile situation in the context of elections in Romania, and the increasing risk of real-world harm and or potential threats to critical democratic institutions, TikTok implemented a temporary search ban on keywords listed above.

97. TikTok does not track specific URLs.

b. A detailed description of the actions that were taken following this Review, including relevant documents attesting such actions.

98. First, as the Commission is aware, Article 8 and Recital 30 DSA expressly state that the DSA does not impose, and nothing in it should be construed as imposing, a general monitoring obligation or a general active fact-finding obligation on providers, nor a general obligation for providers to take proactive measures in relation to illegal content. TikTok understands that nothing in this RFI should be construed as suggesting otherwise.

99. In the context of elections, TikTok has been deploying a range of measures to ensure effective mitigation of risks. These can broadly be divided into: (i) account actions; and (ii) content actions.

Account Actions

100. TikTok has a bespoke reporting tool[37] available to enable the reporting of alleged impersonation accounts (i.e., fake accounts that attempt to impersonate someone else[38]). Moreover, TikTok enables users to report impersonation accounts through TikTok's reporting mechanisms on the Platform[39].

101. In addition, TikTok employs proactive impersonation mitigation measures. These measures were activated in the context of the Romanian elections. They include executing regular sweeps of the Platform, against an identified list of GPPPAs, to identify and remove impersonation accounts.. For the Romanian elections, TikTok used its Romanian GPPPA list (which includes candidates' accounts) as the basis for these sweeps.

102. The sweeps are conducted in accordance with an impersonation policy and sweep detection methodology designed to ensure that impersonation accounts are effectively identified, by identifying, reviewing and enforcing against accounts that bear significant similarities to verified accounts.

103. For the Romanian elections, TikTok began conducting sweeps on 31 October 2024 and continued these efforts throughout the election period. As of 12 December 2024, TikTok had removed 1,996 accounts for violation of its impersonation policies as part of this mitigation strategy and process. Additionally, in this context, TikTok removed 751 accounts for recidivism, belonging to people who had previously been banned from the Platform but continued to create new accounts.

Content Actions

104. In addition to account actions, TikTok takes proactive measures during elections to mitigate against harmful content such as mis/disinformation.

105. In this context, TikTok works with fact-checking partners in identifying potential misinformation in an elections context. These partners actively alert TikTok teams to specific items of content. These are then investigated and enforcement action (including removal) is taken if the content violates TikTok's policies.

---

[37] See https://www.tiktok.com/legal/report/feedback.
[38] See further TikTok's Integrity & Authenticity Community Guidelines at https://www.tiktok.com/community-guidelines/en/integrity-authenticity#6.
[39] See https://support.tiktok.com/en/safety-hc/report-a-problem/report-a-user.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

106. In addition to working with its fact-checking partners, TikTok also conducts internal investigations and sweeps of the Platform. For content, TikTok uses a range of methods to detect violating content, including real-time monitoring of Trust & Safety dashboards, tracking of top reported content with high viewership figures, similarity direction techniques and clustering tools.

107. In the context of the Romanian elections, these methods and measures have led to TikTok removing content identified as violative of relevant Community Guidelines and making certain content FYF-ineligible.[40]

13. Please provide information and data concerning the viral trends during the Romanian elections, including the amount of content flagged for further Manual Review ([41]), the number of views or other engagement signals required to be flagged for further Manual Review, the average and median times taken to reach these threshold values, and the outcomes from the Manual Reviews, including how many pieces of content fell under each outcome and the average and median time taken for Manual Reviewers to reach a decision.

108. TikTok notes that this ***question 13*** refers to the Elections & Integrity Module of TikTok's Year 1 Risk Assessment, which explains that TikTok's Trust and Safety teams undertake targeted sweeps where its human specialists have identified (e.g., on the basis of internal monitoring or external reports) content associated with specific risks. Potentially harmful/violative content identified through these targeted sweeps may then be subject to further manual (i.e., human) review.

109. One factor indicating to TikTok's Trust and Safety teams that content (or specific keywords, hashtags, etc. in content) may require a targeted sweep is if content (including a specific keyword, hashtag, etc.) has gone "*viral*". In this context, "*viral*" refers to a qualitative assessment by TikTok's specialists, not a quantitative metric, and content flagged for manual review is not flagged on the basis of 'virality' alone. TikTok therefore understands that the quantitative metrics the Commission has referred to are not applicable in this case. However, provides following additional information regarding its approach and activities in respect of this mitigation measure:

    i. To support their assessment and identification of "*viral*" trends that might pose a risk to election integrity, TikTok's specialist teams use a number of risk indicators and preventive moderation strategies to identify violative content, including: review of identified threats (e.g., using information through the elections dashboard – please see further TikTok's response to ***question 14*** below); high video views (which, as explained in TikTok's response to ***question 9*** of the Recommender Systems & Elections RFI, is how TikTok detects "*viral*" content); and content that has been widely reported (e.g., through TikTok's Article 16 mechanisms).

    ii. In addition to sweeps, TikTok is developing preventative automated detection models to prevent harmful misinformation (including election misinformation) that may become prevalent across the Platform. In the context of the Romanian elections, TikTok implemented such automated content moderation rules in respect of the following claims:

        "*C. Georgescu (Romanian candidate for Presidency) on Pepsi with nano-chips to control the masses*";

        "*Romanian far-right politician, D. Sosoaca claiming a Civil War is wanted for Romania*";

        "*Putin stopped the pandemic by invading Ukraine*" (this was not specifically related to the Romanian elections, but was identified as a trend in Romania;

        "*Soros is behind protests against C. Georgescu and cancellation of 1st round of elections in Romania*";

[40] https://www.tiktok.com/community-guidelines/en/fyf-standards

[41] Risk assessment 2: integrity & misinformation. Module 1: risks to election and civic integrity, p. 22.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013664

"*C. Georgescu claims people will be hindered from voting for him this weekend*";

"*Lasconi wishing to sell country wealth and strategic state companies*"; and

"*C. Georgescu claims Lasconi and NATO plan to involve Romania in Ukraine war*".

110. These targeted sweeps are distinct from TikTok's identification of "*viral*" content based on 'video views', and the manual review that may follow. These are explained on page 21 of the Elections & Civic Integrity Module Report, and in TikTok's response to ***question 9*** of the Recommender Systems & Elections RFI.

14. The provider of TikTok mentions in its 2023 risk assessment a "*real-time election dashboard*", as a "*critical source of information on evolving trends/threats on the Platform, while also providing a holistic overview of the types and volumes of content relevant to an election. The dashboard enables the core elections team to take informed decisions, including to add new violating keywords or hashtags to keyword lists.*" ([42]). Please provide:

- **A screenshot of the available dashboard tabs and a description of the main features and functionalities;**

111. Under TikTok's Election Integrity Programme, real-time monitoring is enabled by data collection and analysis, presented and visualised via an election dashboard called the "Election Radar". This dashboard provides critical information on evolving trends and threats on the Platform and assists TikTok's Trust and Safety teams by providing direction, information and insights.

112. The "*real-time election dashboard*" referred to in this ***question 14*** is contained within this "Election Radar" dashboard.

113. By way of context, the overall "Election Radar" dashboard includes the following additional features/tabs/functionalities:

    i. **Election Calendar Homepage:** This tracks and outlines global election events and election program core metrics.

    ii. **Election Safety Dashboard:** This presents information on election-related metrics monitoring and analysis.

    iii. **TikTok Ecosystem:** This is a set of toolkits that supports investigation of top-ranking content (i.e., video, hashtag, search, report, etc) and TikTok voice of volume monitoring based on specific keywords.

114. On the real-time election dashboard, this is a dashboard view of the core metrics/trends on content and account activity across the Platform. The dashboard view can reflect aggregated data in 15 minute intervals to allow for increased flexibility and responsiveness. The core tasks enabled by the dashboard view:

    i. tracking of key trust and safety metrics, for certain regions and/or time periods;

    ii. enhanced data visualisation for easier interpretation by safety teams;

    iii. case picking (individual items of content identification) to enable content review visualisation within the wider aggregated data presentation.

115. The real-time election dashboard contains four views (Short Video, Account, LIVE and Photos). These are explained in turn:

[42] Risk assessment 2: integrity & misinformation. Module 1: risks to election and civic integrity, p. 38.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

i. Short Video: This view of the dashboard concerns video content uploaded on the TikTok platform. Data is aggregated to reflect video content that is reported by users as well as enforcement action on video content taken by TikTok.

ii. Account: This view of the dashboard concerns accounts that have been subject to reports and/or moderation.

iii. LIVE: This view of the dashboard concerns livestream content that is subject to reports on the TikTok platform.

iv. Photos: This view of the dashboard concerns photo content that is subject to reports on the TikTok platform.

116. Screenshots of the relevant dashboard are provided with this RFI Response at Annex 001.

- **A list of the datapoints and how they are aggregated for each feature and functionality included in the dashboard;**

117. The list of datapoints pertaining to the real-time election dashboard will be provided as part of TikTok's wider document disclosure for this RFI.

118. The real-time election dashboard gathers data on user signals and TikTok's enforcement activity across the Platforms' features.

119. The core user signal aggregated from a data perspective relates to user reports. For the purposes of the visualisation, these are grouped into three categories: (i) Harassment & Hateful Behaviour; (ii) Integrity & Authenticity (or I&A); and (iii) Other.

120. In addition, TikTok's own enforcement or moderation is captured and aggregated for certain features.

121. This data is captured across regions and time frames. These can be adjusted via search parameters to enable investigation and facilitate accurate mitigation.

- **The trends/threats detected during the period 1 September to 24 November 2024 in the context of the Romanian elections;**

122. Please see TikTok's responses to ***questions 12*** and ***13*** above.

123. The dashboard does not directly identify trends/threats. Instead, TikTok uses the real-time election dashboard to monitor election-related content on the Platform and related metrics. Based on this information, including from the dashboard, TikTok initiates investigations based on anomalies or trends, such as:

i. high numbers of user reports against specific election-related content;

ii. high numbers of views for election-related content; and

iii. general spikes in user reports or moderation action metrics.

124. TikTok's Investigations aim to understand the root cause of the anomalies or trends, including by reviewing the underlying content to assess it for violations of the Community Guidelines or other policies. If violative content is identified due to dashboard monitoring and TikTok's subsequent investigations, TikTok's trust and safety teams transition to conducting further searches for similar content on the Platform using internal investigation tools, including TikTok's similarity detection system, which clusters

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013666

visual signals and keywords. This helps TikTok to ensure the robustness of its detection measures, and facilitates swifter action on violative content and general risk mitigation.

- **Any action taken by the provider of TikTok on the basis of the information gathered through the dashboard, and all internal documents to show how the decisions were taken;**

125. Please see TikTok's responses to ***questions 12*** and ***13*** above.

126. In addition, in certain instances where TikTok has identified significant trends of violative content, TikTok updates its detection models to improve identification of any new similar content on the Platform and automatically enqueue it for moderation.

127. For example, in the context of Romanian elections:

i. On 20 October 2024, TikTok noticed a high number of user reports in the real-time election dashboard against a video featuring Diana Iovanovici-Şoşoacă promoting a dangerous conspiracy theory related to the elections with the narrative "*Americans will rape our girls*". As the video violated TikTok's Community Guidelines, TikTok conducted a follow-up investigation using its similarity detection system to identify a further 32 videos of the same nature and remove them from the Platform before they could cause harm.

ii. On 25 October 2024, TikTok noticed a high number of user reports against a specific user account. This was then investigated and determined as belonging to Codruta Cerva, an individual who was previously banned for spreading misinformation. The investigation identified 3 additional accounts operated by this individual, which TikTok accordingly removed in accordance with the Community Guidelines.

iii. On 22 November 2024, TikTok investigated certain spikes observed via the dashboard. These investigations revealed a cluster of 46 videos containing a number of false claims about George Simion's sexual orientation and a conspiracy theory regarding the dangers of food in supermarkets. These videos were promptly removed.

- **Any document referencing data extracted from the dashboard and including possible follow-up actions.**

128. TikTok does not extract data from the real-time election dashboard. The dashboard serves as a visual aid and information-gathering resource that informs actions TikTok takes on the Platform, including via its other tools and/or measures.

15. Please explain whether the provider of TikTok has identified, analysed and assessed the impact of sending to recipients of the service push notifications from accounts they do not follow on the risk of negative effects on civic discourse and electoral processes. If applicable, please provide all internal documents necessary to show the results of this assessment. In addition, please explain whether the provider of TikTok put in place any measures to mitigate against the risks of this functionality in the context of electoral processes and civic discourse. If yes, please explain in detail these measures, and provide all internal documents discussing their effectiveness.

129. TikTok understands this ***question 15*** to be asking about device-level push notifications (i.e., "*push notifications*" which are received outside the Platform) that recommend content, such as videos, LIVEs or comments, created by accounts that the user receiving the push notification does not follow. TikTok does not understand the question to be asking about, e.g., in-app notifications, or push notifications notifying the user about other users' interactions with that user's own posted content.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013667

130. In this regard, TikTok makes two preliminary points:

    i. First, this ***question 15*** refers to "*the impact of sending to recipients of the service push notifications from accounts they do not follow*".

    ii. Second, all content uploaded to the Platform is assessed at the point of upload by TikTok's automated and/or human content moderation processes, as explained at, e.g., page 86 of TikTok's Year 2 Risk Assessment Report (please also see further TikTok's responses to the Recommender Systems & Elections RFI).

131. TikTok's Articles 34 and 35 DSA Systemic Risk Assessment did not expressly assess the impact of users being sent push notifications from accounts they do not follow on the risk of negative effects on civic discourse and electoral processes (and accordingly does not consider that it has documents responsive to this ***question 15***).

132. However, TikTok has implemented various measures in this area which mitigate against the risks associated with users receiving push notifications related to content from accounts they do not follow. These mitigations include:

    i. measures that are specific to GPPPAs and/or political content; and

    ii. measures that apply to political content more generally, but also serve to mitigate risks in the context of civic discourse and electoral processes.

133. With respect to GPPPA accounts:

    i. In Q3 2024, TikTok analysed the risk of push notifications relating to GPPPAs' content being sent to users.

    ii. Following this assessment, on 17 September 2024, TikTok implemented restrictions on push notifications relating to GPPPAs' content, as a mitigation measure against potential election risks.

    iii. These restrictions apply to all accounts – with the result that neither followers of GPPPAs nor non-followers of GPPPAs receive push notifications for GPPPAs' content. By way of example, TikTok identified (and labelled) the account of Mr Georgescu as a GPPPA. (Please see TikTok's response to ***question 5(a)*** of the Recommender Systems & Elections RFI for information on how TikTok identifies GPPPAs.)

134. With respect to political content:

    i. TikTok has content moderation measures in place which help to ensure that certain content, including certain 'political standpoint content', is not recommended to users via push notifications.

    ii. Such 'political standpoint content' includes, e.g.,:

        - depictions of in-office politicians (in government/party positions);
        - political opinions, satire, commentary, or discussion;
        - political campaigns or appealing for political votes;
        - solicitation for financial support of a party or political cause;
        - advocacy for or against a political party or cause (including political leaders/candidates, political parties/groups, elections, or referenda);
        - content presenting political/military events, or local, state or federal issues (e.g., news, comments, reviews, or documentaries);
        - events related to politics (e.g., foreign visits, meetings between government officials, and campaigns related to a broad range of political issues); and
        - display of political or military symbolism.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013668

135. In addition to these specific measures, general measures that TikTok has implemented to mitigate the risks associated with users receiving push notifications related to content from accounts they do not follow include:

i. Users can decide to disable all, or only certain, push notifications:

- When users open the TikTok app, they are presented with an on-screen message stating: "*TikTok Would Like to Send You Notifications… Notifications may include alerts, sounds and icon badges. These can be configured in Settings*". Users are then presented with two options: "*Don't allow*" and "*Allow*" – with "*Don't allow*" being presented first. Users therefore have a clear choice to opt in or out of receiving push notifications. If they opt in, they can easily change their choice at any time via their device's settings.

- Even if users opt in to receive push notifications generally, they can also use in-app settings to opt out of certain categories of push notifications, including notifications related to content from accounts that they do not follow – e.g., "*Posts from people you may know*" and "*Posts you might like*".

- TikTok's settings[43] also allow users (or a parent/guardian can do so through Family Pairing) to set a "*Push notification schedule*" to turn off push notifications only during a set period of the day.

ii. Additionally, where TikTok's systems identify that users are already connected other than through TikTok (e.g., they are Facebook Friends), they may receive a push notification recommending that they follow each other on TikTok or a video posted by the other user. In such cases, TikTok uses information provided by both users, because they agreed to share either their phone contacts list or friends list on other social media platforms.

iii. Finally, before TikTok launches new features and functionalities related to push notifications, it assesses them through an internal compliance evaluation process. During that assessment, TikTok's cross-functional teams (e.g., Government Relations, Public Relations and Trust and Safety) will raise any identified potential risks.

[43] See further https://support.tiktok.com/en/using-tiktok/messaging-and-notifications/notifications.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013669

Schedule of Annexes

Annex No.	Document
001	Screenshots of real-time election dashboard

The documents listed in this Schedule necessarily reflect matters at the time the documents were created (or last updated, where applicable). As such, some of the information they contain may not fully reflect the current position. Accordingly, please refer to TikTok's responses to the questions of this RFI for the position as at the date of submission of those responses.

Any documents listed in this Schedule have been redacted to remove the following information: (i) personal information; (ii) information that clearly relates solely to countries outside the EEA and is confidential and commercially sensitive, and/or (iii) legally privileged information.

TikTok is producing documents under applicable laws but not waiving privilege claims that may be applicable in any jurisdictions. In the event that any waiver of privilege does nonetheless occur, such waiver is limited to the disclosure of the relevant document(s) to the Commission only and only for the purposes of our response to this RFI.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013670

Exhibit 269

From: [Redacted]tiktok.com]
Sent: 12/17/2024 6:14:53 PM
To: [Redacted]ec.europa.eu]
CC: [Redacted]@tiktok.com]; EuroReg@tiktok.com; CNECT-DSA-REGISTRY@ec.europa.eu; [Redacted] [Redacted]@ec.europa.eu]; [Redacted]ec.europa.eu]; CNECT-F2@ec.europa.eu; [Redacted]ec.europa.eu]; [Redacted] [Redacted]ec.europa.eu]; [Redacted]ec.europa.eu]; [Redacted] [Redacted]ec.europa.eu]
Subject: Re: [External] RE: Romania RFI and in-person meeting

Dear [Redacted]

Yes, this is what I was proposing. Many thanks for the confirmation.
I have confirmed attendees with [Redacted]n this basis.
I look forward to seeing you tomorrow.
Kind regards
[Redacted]

On Tue, 17 Dec 2024 at 15:46, [Redacted]@ec.europa.eu> wrote:

Dear [Redacted]

We can accommodate you. For the avoidance of doubt, in my mind the agenda would then look like this, for your confirmation that I understand your proposal correctly:

1. Retention Order 15:00 – 16:00
2. Prep Croatia and Germany 16:00 – 16:15
3. Summary of the Opening Decision 16:15 – 16:30

Each point would contain its own next steps as the last item.

The timings are what I imagine, please advise if this matches what you had in mind.

From: [Redacted]tiktok.com>
Sent: Tuesday, December 17, 2024 1:10 PM
To: [Redacted]ec.europa.eu>

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Cc: [redacted] tiktok.com>; EuroReg@tiktok.com; CNECT DSA REGISTRY <CNECT-DSA-REGISTRY@ec.europa.eu>; [redacted] [redacted] ec.europa.eu>; [redacted] [redacted] @ec.europa.eu>; CNECT F2 <CNECT-F2@ec.europa.eu>; [redacted] [redacted] @ec.europa.eu>; [redacted] [redacted] ec.europa.eu>; [redacted] @ec.europa.eu>
Subject: Re: [External] RE: Romania RFI and in-person meeting

Dear [redacted]

I am writing regarding tomorrow's meeting and to propose an update to the agenda in light of the opening of the Investigation.

My proposal is that we meet as planned to discuss the Retention Order. We also suggest that we provide an update on our preparation for the Croatia election given this falls during the Christmas period. We can also touch briefly on preparations for the German election, understanding that we will have more opportunity to speak about this in January.

In light of the Investigation, we propose to move the other items to a meeting in the New Year, unless there is something particular that you would like to discuss tomorrow.

If you agree to the above, then my view is that the meeting could be 90 minutes in total with the majority of the time spent on the Retention Order.

Kindly confirm your thoughts on the above and I will then be able to let you know attendees. I very much appreciate your understanding and your flexibility.

Kind regards,

[redacted]

On Mon, 16 Dec 2024 at 17:43, [redacted] @ec.europa.eu> wrote:

Dear [redacted]

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

We can accommodate on the timing to start at 15:00 , though [redacted] will have to leave at 16:00, I will stay for the full meeting. We can accommodate 8 people, please do send the names to [redacted] and [redacted] here in copy.

On the basis of your agenda proposal below, we might prefer to switch the order of Part 1 and Part 2 to allow the right people to be present, but I will confirm tomorrow.

I understand everyone from TikTok's side will be present in person.

[redacted]

From: [redacted]tiktok.com>
Sent: Monday, December 16, 2024 5:49 PM
To: [redacted]@ec.europa.eu>
Cc: [redacted]tiktok.com>; [redacted]tiktok.com; CNECT DSA REGISTRY <CNECT-DSA-REGISTRY@ec.europa.eu>; [redacted] [redacted]ec.europa.eu>; [redacted] [redacted]ec.europa.eu>; CNECT F2 <CNECT-F2@ec.europa.eu>; [redacted] [redacted]ec.europa.eu>; [redacted] [redacted]ec.europa.eu>; [redacted] [redacted]@ec.europa.eu>
Subject: Re: [External] RE: Romania RFI and in-person meeting

Dear [redacted]

a quick correction to the email below is that there would be 8 attendees for the first part. If that is difficult to accommodate, please do let me know.

I look forward to hearing from you.

Best

[redacted]

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

On Mon, 16 Dec 2024 at 11:39, Redacted tiktok.com> wrote:

Dear Redacted

Thank you for your email on Friday, agreeing to an in-person meeting with TikTok on Wednesday 18 December. I set out some follow up comments below.

Time proposed

First, in terms of the proposed time, I wanted to understand if there is a chance to move the meeting to start at 3pm. The reason for this is that TikTok has a meeting with BNetzA on the German election at this time and we would ideally like one of our team members to be able to attend that meeting as well as the meeting with yourselves. I appreciate the challenges of logistics and diary coordination at your end, so if moving the meeting is not possible, then we will proceed as planned at 2pm. I apologise for not mentioning this last week.

Agenda

In terms of topics for the meeting we propose that we split this into two parts:

Part 1:

- The structure of our trust and safety/content moderation teams working on election integrity

- How we deal with suspected CIOs and inauthentic behaviour

- Preparation for the Croatian and German elections.

Part 2: the Retention Order

If there are other topics that you think would be useful, please let me know.

Attendees

In terms of attendees we would like to have 9 people for the first part and 4 for the second part. Please confirm that you can accommodate this, and I will then provide the list of names. We would like our external counsel to attend the second part. We do not currently plan for anyone to be dialling in.

I want to say thank you once again for your willingness to meet this coming week.

Kind regards,
Redacted

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

On Fri, 13 Dec 2024 at 16:42, Redacted @ec.europa.eu> wrote:

Dear Redacted

I can confirm Wednesday 14:00 for an in-person meeting in the Commission's premises in Brussels for a meeting as requested below.

I can also confirm that participants on our side will include at least Redacted and myself. We will confirm additional participants as soon as we know the additional agenda items you mention below.

I copy in Redacted personal assistant and Redacted from my team to facilitate access to the building for your team + confirmation on the need for a hybrid connection and the necessary details.

We are aiming for 90 minutes but please do let us know if this fits in the agenda items that you have in mind.

Kind regards

PS: When you submit the RFI documents later today, please do keep me and Redacted in copy in addition to the DSA Registry.

From: Redacted @ec.europa.eu>
Sent: Thursday, December 12, 2024 10:01 PM
To: Redacted tiktok.com>
Cc: Redacted tiktok.com>; Redacted tiktok.com; CNECT DSA REGISTRY <CNECT-DSA-REGISTRY@ec.europa.eu>; Redacted Redacted ec.europa.eu>; Redacted Redacted ec.europa.eu>; CNECT F2 <CNECT-F2@ec.europa.eu>
Subject: Re: Romania RFI and in-person meeting

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Dear Redacted

Thank you.

We can grant you an extension for the RFI dated 6 December to Monday 16 December 10am Brussels time.

We would also very much welcome an in-person meeting on Wednesday afternoon in Brussels.

I will revert tomorrow with an exact timing proposal and the attendance on our side.

Redacted

From: Redacted tiktok.com>
Date: Thursday, 12 December 2024 at 19:54:16
To: Redacted ec.europa.eu>
Cc: Redacted @tiktok.com>, "Redacted tiktok.com" <EuroReg@tiktok.com>, "CNECT DSA REGISTRY" <CNECT-DSA-REGISTRY@ec.europa.eu>
Subject: Romania RFI and in-person meeting

Dear Redacted

thank you for your time this morning.

As discussed, we have been working tirelessly on the 3 RFIs due tomorrow and the Preservation Order. We would be very grateful if you would kindly extend the deadline for the provision of the responses to Section 1.2 in the RFI formally served on 6 December.

We think it would be very helpful to have an in-person meeting with yourself, Redacted and Redacted on Wednesday 18 December in Brussels. This would be with two objectives: (1) to discuss the compliance with the Preservation Order (details of which you will receive by the end of today); and (2) to discuss some more specific matters around Romania and the forthcoming elections including information about our content

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

moderation teams and our readiness for the Croatia election. I am also discussing internally other potential agenda items that may be of particular interest to you and I will follow up with you next week. I would be grateful if you could confirm your and Redacted availability for such a meeting and then we can pick up on the full list of attendees at the start of next week.

Kind regards

Redacted

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 270

From: COP Rapid Response System [cop-rrs-elections@tiktok.com]
Sent: 12/20/2024 6:32:07 PM
To: Redacted funky.ong]
CC: Redacted ec.europa.eu]; Redacted gate-ai.eu]; Redacted funky.ong]; CNECT-COP-TASK-FORCE@ec.europa.eu; Redacted vosteurope.org]; Redacted Redacted tiktok.com]; Redacted tiktok.com]
Subject: Re: [External] RRS Romania

Hi Redacted

As per our conversation during the last meeting, we would like to follow up on the RRS report received on 3.12.2024.

- On top of previously communicated community guidelines violations, access to further **66 videos/photo posts** were blocked in Romania for violating local law.

Kind regards,
TikTok T&S Team

From: "COP Rapid Response System"<cop-rrs-elections@tiktok.com>
Date: Thu, Dec 12, 2024, 13:26
Subject: Re: [External] RRS Romania
To: Redacted @funky.ong>
Cc: Redacted @ec.europa.eu>, Redacted gate-ai.eu>, Redacted Redacted @funky.ong>, "cnect-cop-task-force@ec.europa.eu"<CNECT-COP-TASK-FORCE@ec.europa.eu>, Redacted @vosteurope.org>, Redacted Redacted @tiktok.com>, Redacted @tiktok.com>

Hi Redacted

Thanks for your email.
We would like to follow up on the RRS report received on 3.12.2024, discussed during our first sub-group meeting.

- 94 links received referring to 94 video/photo posts; Among the 94 video/photo posts reported:
  - 12 videos/photo posts in violation of Hate speech, Misinformation, Civic and Election Integrity policies.
  - 2 videos/photo posts were already removed or permanently deleted by the user by the time they reached us

We are also happy to bring to your attention two blogposts recently shared, including Continuing to protect the integrity of TikTok during Romanian elections and How TikTok counters deceptive behaviour.

Thanks,
TikTok T&S Team

From: Redacted funky.ong>
Date: Tue, Dec 3, 2024, 14:37
Subject: [External] RRS Romania
To: <cop-rrs-elections@tiktok.com>
Cc: Redacted cc.europa.eu>, Redacted gate-ai.eu>, Redacted Redacted funky.ong>, "cnect-cop-task-force@cc.europa.eu"<CNECT-COP-TASK-FORCE@cc.europa.eu>, Redacted democracy-reporting.org>, Redacted @democracy-reporting.org", Redacted @democracy-reporting.org>,

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted vosteurope.org" Redacted @vosteurope.org>,
Redacted alliance4europe.eu" Redacted @alliance4europe.eu>

Good evening,

Please see in the spreadsheet some flags identified by my colleagues on their elections monitoring in RO:

Redacted

Thank you,

Redacted

Funky Citizens
www.funky.ong
Tel: Redacted
E-mail: Redacted @funky.ong

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 271

From: CNECT-COP-TASK-FORCE@ec.europa.eu [CNECT-COP-TASK-FORCE@ec.europa.eu]
Sent: 2/3/2025 10:41:12 AM
To: Redacted tiktok.com]; Redacted
CC: Redacted tiktok.com]; Redacted tiktok.com; Redacted
[a Redacted tiktok.com]; Redacted meta.com]; Redacted microsoft.com];
Redacted linkedin.com]; Redacted google.com]; Redacted
Redacted @google.com]; Redacted google.com; Redacted meta.com]; Redacted
[Redacted ec.europa.eu]; Redacted @ec.europa.eu]; Redacted
Redacted @ec.europa.eu]; Redacted ec.europa.eu]; Redacted
Redacted ec.europa.eu]
Subject: [External] RE: CoP - German Federal Election - RRS

Dear Redacted dear all,

Thank you for your email, we acknowledge receipt and will take the necessary steps to operationalize the RRS today.

Two small clarifications from our side:

- We agree to hold meeting on a need-basis only, which means that – while we aim to keep the meetings cadence at once a week – we might need to increase if the situation warrants it.
- As for the past exercises, the relevant EDMO hub (GADMO – covering Germany and Austria) will be participating in this iteration.

We will contact the WG Members to get the full list of participants and will share with you as soon as possible.

If you could also share all the relevant email addresses/forms – including one to be contacted in case of issues with internal tools.

Thanking you for the collaboration
Kind regards,
Redacted for the team

From: Redacted tiktok.com>
Sent: Friday, January 31, 2025 6:06 PM
To: CNECT COP TASK FORCE <CNECT-COP-TASK-FORCE@ec.europa.eu>; STUMP Krisztina (CNECT)
Redacted @ec.europa.eu>
Cc: Redacted tiktok.com>; Redacted tiktok.com; Redacted
< Redacted tiktok.com>; Redacted meta.com>; Redacted microsoft.com>; Redacted
Redacted @linkedin.com>; Redacted google.com> Redacted
< Redacted google.com>; Redacted google.com; Redacted meta.com>
Subject: CoP - German Federal Election - RRS

Dear Redacted and team,

I hope this finds you well.

Having aligned internally with our respective teams and among all VLOPs signatories, we can confirm our participation to the RRS for the upcoming German Federal Elections. I'm also including a detailed follow up on the requirements discussed during our last sub-group meeting earlier this week:

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

- **RRS Activation Period** - as per proposal, we agree on 3 weeks before and up to 1 week after the election.
- **Language/s in scope** - we agreed that the scope of RRS should be primarily focused on German and English. High-risk Arabic, Turkish, or Russian content—posing a serious or systemic threat to election integrity—may also be submitted, along with detailed reasoning, fact-checking verification where possible, and language information.
- **Reporting guidelines** - For all flags, as we have agreed before, we would appreciate having the following details:
    - Name of flagging organizationTime of notification (including time of notification via regular tools)
    - Platform concerned
    - URL to the content
    - Language of the content
    - Reason for the flag (including if possible reference to relevant platform policies)

As per usual, in order to activate the RRS, we would need the **email list of CSO/Fact-Checking organizations that will participate**, as some VLOPs will need to onboard them to their respective intake form/tools (e.g. META, TikTok). Each platform will then communicate back the designated channel for the RRS.

Lastly, as per previous RRS activation, we ask that calls be held on a need-basis only (i.e. If there are flags and with those companies that received said flags), not more than once a week.

Many thanks,

Redacted

--

Redacted

T&S Integrity & Authenticity Product Lead | Dublin

Redacted

Redacted Redacted .com

Follow us on

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 272

From: Redacted@tiktok.com]
Sent: 2/20/2025 11:20:20 AM
To: Redacted ec.europa.eu]
CC: Redacted tiktok.com]; Redacted ec.europa.eu]; Redacted @ec.europa.eu]; CNECT-DSA-TikTok-Case-Team@ec.europa.eu; CNECT-Redacted ec.europa.eu; Redacted cnam.ie]; Redacted ec.europa.eu]; EUROREG TikTok [euroreg@tiktok.com]
Subject: Re: [External] Confirmation of Product Team meeting date and topics to cover
Attachments: Confidential EC TikTok Meeting on Elections Investigation 20 Feb 2025.pdf

Dear Redacted and DSA Team,

As promised, please find attached slides which we intend to present during our meeting with you today.

The attached documents contain confidential information and business secrets within the meaning of Article 6 of Commission Implementing Regulation (EU) 2023/1201. Such information is supplied to the Commission exclusively in relation to TikTok's compliance with Articles 34 and 35 of Regulation (EU) 2022/2065 (the DSA), and its publication or disclosure would harm the legitimate commercial interests of TikTok.

We look forward to seeing you shortly.

Kind regards,
Redacted

On Wed, Feb 19, 2025 at 4:07 PM Redacted tiktok.com> wrote:
Thank you very much Redacted

Kind regards,
Redacted

On Wed, Feb 19, 2025 at 4:02 PM Redacted ec.europa.eu> wrote:

Dear Redacted

Thanks for confirming.

Please find the WebEx details below this email.

Redacted

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

***

CNECT-WEBCONFERENCE-03 is inviting you to a scheduled Webex meeting.

Thursday, February 20, 2025

2:00 PM | (UTC+01:00) Brussels, Copenhagen, Madrid, Paris | 2 hrs

Join meeting

More ways to join:

Join from the meeting link

https://[redacted]

5d65a8322c

Join by meeting number

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Meeting number (access code): [redacted]

Meeting password: [redacted] ([redacted] when dialing from a phone)

Tap to join from a mobile device (attendees only)

+[redacted]# Belgium Toll

+[redacted]# Luxembourg Toll

Some mobile devices may ask attendees to enter a numeric password.

Join by phone

+[redacted] Belgium Toll

+[redacted] Luxembourg Toll

Global call-in numbers

Join from a video system or application

Dial [redacted]@ecconf.webex.com

You can also dial [redacted] and enter your meeting number.

Need help? Go to https://help.webex.com

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

From: Redacted tiktok.com>
Sent: Wednesday, February 19, 2025 3:49 PM
To: Redacted @ec.europa.eu>
Cc: Redacted tiktok.com>; Redacted Redacted ec.europa.eu>; Redacted @ec.europa.eu>; CNECT DSA TikTok Case Team <CNECT-DSA-TikTok-Case-Team@ec.europa.eu>; CNECT F2 <CNECT-F2@ec.europa.eu>; Redacted <Redacted cnam.ie>; Redacted ec.europa.eu>
Subject: Re: [External] Confirmation of Product Team meeting date and topics to cover

Thank you Redacted The following TikTok team members will join remotely:

>From Dublin:

Redacted

>From London:

Redacted

Could you perhaps send me the Webex link that I can share out to my colleagues?

Thank you for making the room available from 1.30pm. I will be able to share slides with you tomorrow morning.

Kind regards,

Redacted

On Wed, Feb 19, 2025 at 1:07 PM Redacted ec.europa.eu> wrote:

> Dear Redacted
>
> Thank you for confirming the additional attendees on your side. Could you let me know whether they will be joining in-person? In that case, we will make sure they are registered in our entry pass system too.
>
> We've booked the room at 13.30 and you're welcome to use if from that time to get set up.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

It should normally be possible to project on the screen through a MacBook (and if not, there's the back-up option of joining the WebEx meeting and sharing the screen there). In any event, we'd appreciate it if you could share the slides with us before the meeting tomorrow, also in case any technical issues arise.

Best regards,

Redacted

From: Redacted tiktok.com>
Sent: Wednesday, February 19, 2025 11:22 AM
To: Redacted @ec.europa.eu>
Cc: Redacted tiktok.com>; Redacted Redacted @ec.europa.eu>; Redacted ec.europa.eu>; CNECT DSA TikTok Case Team <CNECT-DSA-TikTok-Case-Team@ec.europa.eu>; CNECT F2 <CNECT-F2@ec.europa.eu>; Redacted <Redacted cnam.ie>; Redacted ec.europa.eu>
Subject: Re: [External] Confirmation of Product Team meeting date and topics to cover

Dear Redacted

Thank you very much for confirming the Commission's attendees.

We will have additional 3 attendees on our side:

>From Dublin:

Redacted

>From London:

Redacted

Could I please also ask that we have access to the room 15mn before the start of the meeting? This would allow us to get set up and ensure that we can project our presentation. We will be projecting via Macbook. Let me know whether that could work on your end.

Thank you very much, and we look forward to meeting with you and the DSA team tomorrow.

Kind regards,
Redacted

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

On Tue, Feb 18, 2025 at 9:11 AM Redacted @ec.europa.eu> wrote:

Dear Redacted

Thank you for confirming the list of attendees on TikTok's side. You should have received a WebEx link for remote attendees.

On our side the following people will join:

Redacted

We have also asked our CnaM colleagues, who are associated to the investigation, to join the meeting remotely Redacted is in copy of this email).

Best regards,

Redacted

From: Redacted tiktok.com>
Sent: Friday, February 14, 2025 6:14 PM
To: Redacted ec.europa.eu>
Cc: Redacted tiktok.com> Redacted Redacted ec.europa.eu>; Redacted ec.europa.eu>; CNECT DSA TikTok Case Team <CNECT-DSA-TikTok-Case-Team@ec.europa.eu>; CNECT F2 <CNECT-F2@ec.europa.eu>
Subject: Re: [External] Confirmation of Product Team meeting date and topics to cover

Dear Redacted

Thank you again very much for sending us this set of areas which the EC would like to cover.

I confirm that the date works for our teams, and that the call will be hybrid. Most of our team members are based in Dublin, and given the focus on the upcoming German elections, they unfortunately, cannot travel next week. However: myself, Redacted and our regulatory counsel specialised on ads, will join you in Brussels.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Here is the full list of attendees:

Dublin attendees remotely:

Redacted

(I am waiting to confirm the attendance of one other colleague - which I will communicate next week).

Brussels attendees in person:

Redacted

Would you please be able to send us a Webex Link for our colleagues who are joining from Dublin to dial into the meeting?

We can address the points which you have suggested to cover, however, we might change the order of these items. Additionally, TikTok can share with you some of the latest developments with respect to the upcoming German elections.

Please let me know if you have any questions.

Kind regards,
Redacted

On Wed, Feb 12, 2025 at 11:17 AM Redacted tiktok.com> wrote:

Thank you Redacted for sending us the list of topics you would like to be addressed at the meeting and for our call this morning.

We will be discussing attendees internally based on the specific topics. I will get back to you shortly to confirm the date and attendees on our end.

Thank you for accommodating us, we appreciate it.

Kind regards,

Redacted

On Wed, Feb 12, 2025 at 11:00 AM Redacted cc.europa.eu> wrote:

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Dear Redacted

Thank you for the call earlier this morning.

As discussed, we are coming back to you to confirm the new date of the product team meeting: **Thursday, 20 February, 14.00-16.00**.

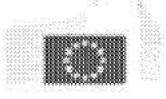
The areas we would like to cover in the meeting are:

1. Fake engagement & deceptive behaviour policies & enforcement
2. Potential ineffectiveness of model detection for viral content
3. The 'Promote' feature
4. Boosting engagement by TikTok on FYF
5. Potential ineffective mitigation measures for search recommendations
6. Review of political ads and enforcement of branded political content policies

Can you please **confirm if the above date works for you** and **the names and email addresses of the participants** on your side?
Based on the topics above, please do send **a proposal for the agenda** as well.

Thank you and have a great rest of your day!
Best regards,
Redacted for the team

Redacted



European Commission

DG CNECT - Directorate-General for Communications Networks, Content and Technology

Platforms Policy and Enforcement

Unit F2 – Digital Services
Rue de la Loi 51, L-51

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

B-1049 Brussels, Belgium
Redacted ec.europa.eu

DIGITAL STRATEGY - Privacy statement related to activities undertaken under the Digital Services Act (europa.eu)

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 273

Message

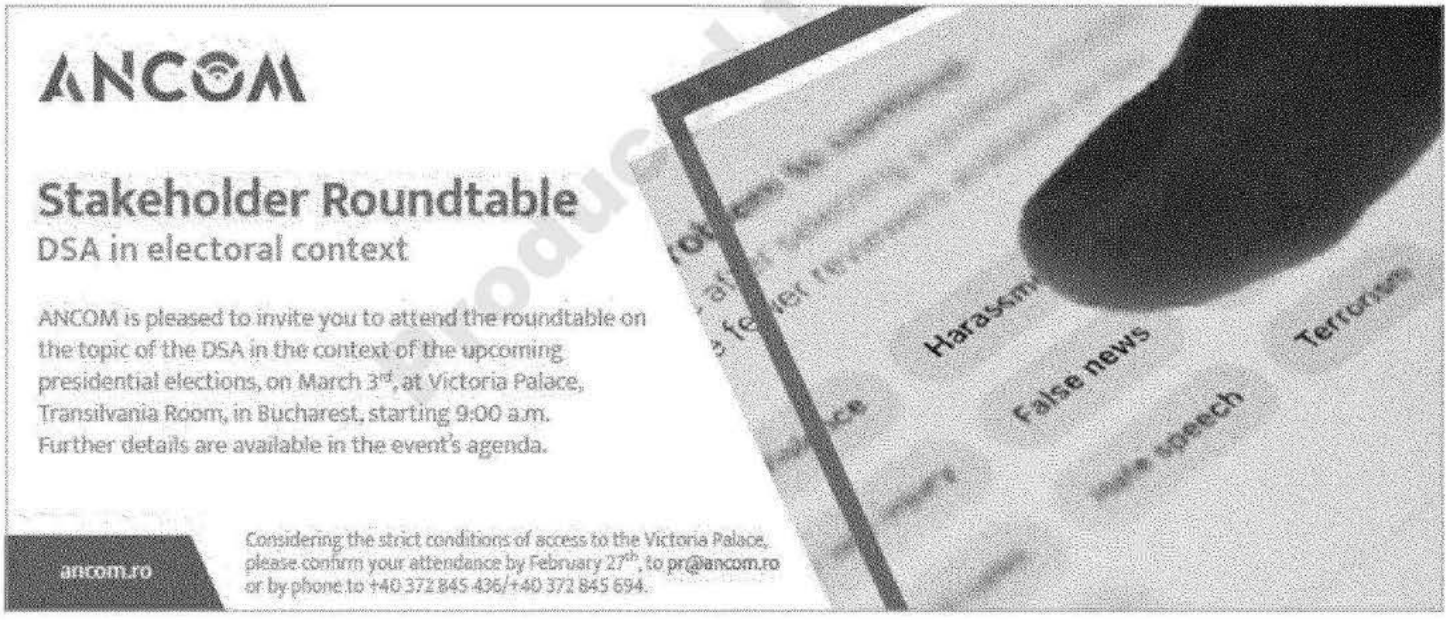
From: Meta DSA Compliance Office [[redacted]@meta.com]
Sent: 2/24/2025 9:44:37 AM
To: DSA Inbox Group [[redacted]@meta.com]
CC: Meta DSA Compliance Office [[redacted]@meta.com]
Subject: Fw: Invitation - Stakeholder Roundtable on DSA electoral context
Attachments: Agenda_roundtable_RO elections.pdf

Hi All, sharing for Viz. We received the invite and agenda for the Romanian Election Roundtable next Monday 3 March.

Thanks,
[redacted]

From: ANCOM PR[redacted]@ancom.ro>
Sent: Monday 24 February 2025 3:25 pm
To: ANCOM PR[redacted]@ancom.ro>
Subject: Invitation - Stakeholder Roundtable on DSA electoral context

This Message Is From an External Sender



Best regards,
DSC Romania

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 274

ANCOM | Autoritatea Națională pentru Administrare și Reglementare în Comunicații

Str. Delea Nouă nr.2, Sector 3, Bucureşti, CP 030925, România
Tel: 0372 845 600 / 0372 845 601, Fax: 0372 845 602
ancom@ancom.ro www.ancom.ro

Draft agenda
Bilateral meeting ANCOM & TikTok
Date: 27 February 2025
Time: 11:00-12:30, Bucharest time
Venue: online.

Agenda

1. Opening Remarks
   - Meeting objectives and importance of collaboration for election integrity.
   - Overview of the obligations of VLOPs/VLOSEs under the Digital Services Act (DSA).
2. Systemic Risks, Disinformation & Election Integrity
   - Detection and mitigation of coordinated inauthentic behaviour (bots, troll networks, fake accounts).
   - Mechanisms for identifying and combating electoral disinformation.
   - Reporting channels, moderation teams, response times, and available resources.
   - Measures taken by VLOPs/VLOSEs to address election-related risks and ensure compliance with the European Commission's Guidelines on systemic risk mitigation.
3. Closing Remarks
   - Summary of key points and next steps.
   - Follow-up actions and timeline for implementation.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_019977

Exhibit 275

ANCOM | National Authority for Management and Regulation in Communications of Romania

2 Delea Noua Street, Bucharest 3, 030925 Romania
Tel: +40 372 845 400 / 0372 845 454 Fax: +40 372 845 402
ancom@ancom.ro www.ancom.ro

Roundtable - 2025 Romanian Presidential Elections

3 March 2025, Bucharest
Victoria Palace, 1 Victoria Square, Sector 1
Transilvania Room, 1st Floor
- Draft agenda -

Registration, welcome coffee		09.00
Opening session	*Welcome speech* **Valeriu ZGONEA**, President of ANCOM *Opening remarks* **Representative of** Romanian Government *(TBC)* **Representative of** DG CONNECT, European Commission	09.30 - 09.40
Main session	*Moderator/Host:* **ANCOM** *Participants:* **Representatives of CSOs** **Representatives of academia** **Representatives of VLOPs/VLOSEs** **Representatives of national authorities (e.g. National Authority for Management and Regulation in Communications, Permanent Electoral Authority, Central Electoral Bureau, National Audiovisual Council, Romanian National Cyber Security Directorate, Ministry of Home Affairs, Romanian Police, other law enforcement units, Ministry of Foreign Affairs)** **Representatives of European Commission**	09.40 - 12.00
	Discussions will focus on the obligations under the Digital Services Act (DSA) of very large online platforms and search engines (VLOPs/VLOSEs) in the run-up to and during elections, measures that VLOPs/VLOSEs should take when/if a breach of the DSA arises in relation to the elections in May 2025, cooperation channels between VLOPs/VLOSEs and the relevant national authorities etc. Furthermore, building on the experience of the past 2024 elections, the discussions will touch upon issues such as mitigation measures envisaged or already put in place by VLOPs/VLOSEs, in line with the recommendations included in the European Commission's Guidelines on the mitigation of systemic risks for electoral processes, reporting channels available, moderation teams used by the platforms, time of response, resources (both human and technological) allocated during elections, identification of inauthentic behaviour, fact-checking programmes for Romania (if any), localised media literacy campaigns (if any), access to platforms' data by NGOs, rapid response system (RRS), electoral labelling and badging for political accounts, mechanism for identifying and combating electoral disinformation, management of controversial electoral content, limitations and restrictions on political ad targeting (with focus on unethical payments for political content as commercial ads), flags for coordinated dis/misinformation campaign, existence of standardized reporting formats, examples of clear violations of terms&conditions, suggestions for improvement from VLOPS/NGOs etc.	
Q&A session		12.00 - 12.30
Closing remarks	**Valeriu ZGONEA**, President of ANCOM, Romania	

Provided to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

META-119HJC-0001094

Exhibit 276

From: Sabajo, Manisha <[redacted]@acm.nl>
Sent: Wednesday, 3 September 2025 12:09
To: Korteweg, David <[redacted]@acm.nl>; Visser, Ernst <[redacted]@acm.nl>
Cc: Burkunk, Merel <[redacted]@acm.nl>; Megen, Stijn van <[redacted]@acm.nl>; Looze, Joep <[redacted]@acm.nl>
Subject: [EXTERNAL] Update DSA roundtable discussion on elections in the Netherlands

Dear all,

A brief update about our upcoming DSA Election roundtable that you will join on September 15th.

Please find below the organisations that currently confirmed their attendance to the roundtable discussion. We are still waiting for a few additional organisations to confirm their participation. As part of our commitment to transparency, we will make the list of participating organisations publicly available. This list will only include organization names, not the names of the individual participants.

For those planning to come by car, please let me know in advance so I can reserve a parking spot. Kindly also remember to bring a valid ID for security purposes when entering the venue.

We hope the discussion will be both engaging and productive. If you have any questions in the meantime, please don't hesitate to reach out to me directly.

Alphabet
Autoriteit Persoonsgegevens (Dutch Data Protection Authority)
Bits of Freedom

MICROSOFT CONFIDENTIAL

Coimisiún na Meán (Ireland's media regulator)
Council of State (Raad van State)
DG CNECT and DG JUST (European Commission)
Discord
Dutch Media Authority (Commissariat van de Media)
INACH
Institute for Information Law (IViR) (UvA)
Justice for prosperity
Linkedin
Maastricht University
Media Studies (University of Amsterdam)
Meta
Microsoft
Ministry of the Interior and Kingdom Relations
Nieuwscheckers
Post X Society / Democratie Monitor
Radboud University
Snapchat
Tiktok
Wikimedia Foundation
Wikimedia Nederland
X

Kind regards,

MICROSOFT CONFIDENTIAL

Manisha Sabajo

Senior Regulator Digital Platforms

Telecommunications, Transport and Postal Services Department

Authority for Consumers & Markets

M: +[redacted]

www.acm.nl/en

www.acmconsuwijzer.nl

Muzenstraat 41
2511 WB The Hague
The Netherlands

P.O. Box 16326
2500 BH The Hague
The Netherlands

Markets that work well for people and businesses

* *This e-mail is intended for the exclusive use of the recipient and may contain confidential information. Read here the disclaimer that applies to this message.*

Exhibit 277

Authority for Consumers & Markets

Muzenstraat 41
2511 WB The Hague
+[redacted]
www.acm.nl

Roundtable on Elections in the Context of the Digital Services Act - September 15, 202

Part I: online platforms and government bodies only		
Time	Themes	Subjects
12:30 – 13:00	Walk-in	Sandwiches will be available for participants who may not have the opportunity to have lunch beforehand.
13:00 – 13:05	Welcome by David Korteweg (ACM)	- Agenda - Houserules
13:05 – 13:25	Presentation by the European Commission	- The role of the DSA in the electoral process - Election Guidelines - Code of Practice on Disinformation - TTPA
13:25 – 14:00	Overview of Dutch context	- Landscape of Dutch election process and (government) actors – *Ministry of the Interior and Kingdom Relations* - TTPA – *Dutch Data Protection Authority & Dutch Media Authority*
14:00 – 14:15	Measures for politicians/political parties	- Threats against politicians - Contact point/action perspective for political parties/politicians - AI generated information about candidates and political parties
14:15 – 14:30	Break	
Part II: online platforms, government bodies, civil society, academics		
14:30 – 14:35	Opening by Manon Leijten (Board Member of the ACM)	
14:35 – 15:05	Synthetic content: political issues and election information	- *Introduction by external speaker* - Synthetic content about political issues - Chatbots/AI: responses about parties, politicians and elections
15:05 – 15:35	Recommender systems	- *Introduction by external speaker* - How content is prioritized - System abuse
15:35 – 16:05	Access to publicly available data for researchers	- *Introduction by external speaker* - Access pursuant to Art. 40(12) DSA: experiences and obstacles
16:05 – 16:20	Break	
16:20 – 16:50	Virality	- *Introduction by external speaker* - Timely detection and intervention of harmful content - Cross-platform distribution
16:50 – 17:20	Open discussion	- What structural improvements are needed before the elections? - What is needed from the ACM/DSCs? - Collaboration between researchers and platforms
17:20	Networking drinks	

MICROSOFT CONFIDENTIAL

Exhibit 278

NL Elections Roundtable September 15th, 2025

SUMMARY

On September 15, the Netherlands Authority for Consumers and Markets (ACM) organized a roundtable discussion on the responsibilities of online platforms under the European Digital Services Act (DSA) during elections. During the meeting, major online platforms, civil society organizations, researchers, and relevant authorities discussed this topic. The goal of the meeting was to exchange knowledge about the relevant legislation, each party's role, and the experiences of academic experts. The roundtable discussion is part of a standard roadmap developed at the European level. Apart from the ACM and invited VLOPs and VLOSEs, attendees included the European Commission, the Dutch government, civil society organizations and academia.

Meeting summary

The meeting focused on four issues:

- deepfakes created on platforms like Instagram, X and Tiktok
- promotion of hatespeech and misinformation (incl harassment of politicians/journalists)
- fact-checking and community notes lack of effectiveness in virality
- Slowness in content removal rate when requested to do so

The EU Commission briefing on the Digital Services Act (DSA) and elections focused on how the DSA helps safeguard electoral integrity by requiring platforms to mitigate systemic risks like disinformation and threats to civil discourse. Over the past two years, the Commission has developed guidelines and an Election Toolkit to support compliance, especially for Very Large Online Platforms (VLOPs). National authorities retain control over elections, but collaboration is key—fact-checkers and EDMO hubs play a role in rapid response systems, while a new political advertising transparency regulation (effective October 10th, but EU guidelines not in place yet) will require sponsor disclosure without regulating ad-content. Discussions also covered harassment of journalists and politicians, the impact of AI-generated content and deepfakes, challenges in accessing platform data for research, and the limited effectiveness of current tools like fact-checking and community notes in curbing viral misinformation. Finally, an interesting mention was made on the interlink between the 'Dutchification of politics' (the Dutch political system can lead to political fragmentation, electoral volatility and the rise of niche parties. In the 2023 elections fx 26 parties participated, of which 15 parties ended up in parliament) and the expected increased use of AI chatbots to differ between a variety of political parties on the left, or right end of the spectrum.

ACM published a high-level summary of the Roundtable on their website: Verslag rondetafelgesprek over verantwoordelijkheden online platforms in verkiezingstijd | ACM

MICROSOFT CONFIDENTIAL MSFT_JUD_000004172

FURTHER DETAILED NOTES

EU commission intro on DSA

- **Focus**: How the DSA interacts with elections, systemic risk mitigation, and civil discourse.
- **Key Tools**:
  - **Election Toolkit**: Developed by the Commission and Digital Services Board to guide relevant actors.
  - **Rapid Response System**: Enables fact-checkers to flag harmful content to platforms.
  - **EDMO**: Network of fact-checkers across EU (e.g., NL hub).

Political Advertising Regulation

- **New Regulation**: Enters into force on **Oct 10**.
- **Goals**: Transparency in political ads (e.g., sponsor details), resilience of political systems. On prior elections they were in touch with SOME platforms to ask them to review message -- election flagger status.
- **Scope**: Online and offline ads; does **not** regulate ad content.
- **Transparency Labels**: Required since **July 9**.

NL Dutch National Context

- **Ministry of Interior**: Legal framework, voting process, and election interference.
- **Election Flagger**: Informal status used to request content review from platforms.
- **Dutch Media Authority**: Monitoring and notification role.
- **Dutch Data Protection Authority**: Supervises GDPR and recommender system transparency.

Harassment & Threats to Politicians and Journalists

- **Findings**:
  - 80% of female journalists received threats; 25% reported mental health impact.
  - 53% of female politicians experienced harassment.
- **Questions Raised**:
  - Do these groups need special escalation channels?
  - Can these be shared with candidates?

AI-Generated Content & Deepfakes

- **Valentina Golunova (Univ. Maastricht)**:
  - Deepfakes can violate GDPR and electoral integrity.
  - Governed by EU law, national law, and platform policies.
  - DSA Articles: 9, 16, 22, 34–37; AI Act: Article 50(2)(4) (labeling by Aug 2026).
- **Platform Examples**:
  - **X**: Prohibits deepfakes but doesn't remove them.
  - **Meta**: Oversight board recommends removal, not enforced.
  - **TikTok**: Deepfakes not labeled.

Recommender Systems & Content Moderation

- **Concerns**:

- Promotion of harmful content (e.g., fascism, stereotypes).
- Lack of user control over feed (e.g., chronological order).

- **Speakers**: Bits of Freedom, Post X Society, TikTok, Univ. of Amsterdam.
- **Need for Collaboration**: CSOs, academia, and governments must work together.

Data Access for Researchers

- **Article 40(12)**: Access to public data.
- **Article 40(4)**: Future access to non-public data.
- **Challenges**: Platform approval rates low; Meta's CrowdTangle shutdown impact.
- **Outlook**: Need for timely, accurate, and useful data access.

Virality & Misinformation

- **Univ. Amsterdam**:
  - Fact-checkers vs. Community Notes: Both ineffective at reducing virality.
  - Community Notes: Only 10% published; not active in EU.
  - Wikimedia: Believes in user self-regulation.

Monitoring & Enforcement

- **INACH CSO**:
  - Tracks platform response to harmful content.
  - Compares removal rates between regular users and trusted partners.
  - Ongoing monitoring since 2016.

Open Questions & Challenges

- Experience with DSA Transparency Database.
- Role of platforms in countering hate speech and threats.
- Journalist protection and cybersecurity training.
- Timing and readiness for deepfake interventions.

Participants:

Alphabet
Autoriteit Persoonsgegevens (Dutch Data Protection Authority)
Bits of Freedom
Coimisiún na Meán (Ireland's media regulator)
Council of State (Raad van State)
DG CNECT and DG JUST (European Commission)
Discord
Dutch Media Authority (Commissariat van de Media)
INACH
Institute for Information Law (IViR) (UvA)
Justice for prosperity
Linkedin
Maastricht University
Media Studies (University of Amsterdam)
Meta
Microsoft
Ministry of the Interior and Kingdom Relations
Nieuwscheckers
Post X Society / Democratie Monitor
Radboud University
Snapchat
Tiktok
Wikimedia Foundation
Wikimedia Nederland
X

Exhibit 279

Tab 1

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Coimisiún na Meán (CnaM) Irish Presidential Elections Roundtable - 24 September 2025 - Detailed Meeting Readout

Purpose: Coimisiún na Meán, Ireland's Digital Services Coordinator, to invite you to attend a Digital Services Act – Election Roundtable, which will involve online platforms and other State Agencies and regulators, with participation from the European Commission. As Ireland's Digital Services Coordinator, Coimisiún na Meán is responsible for the application of the EU Digital Services Act in Ireland.

Location: Hybrid - Clayton Hotel Burlington Road Dublin and VC.

Date / Time: Wednesday 24th September 2025, 10:00-13:00.

Roundtable Chair: Stephanie Comey, Director of Media Literacy and User Education at Coimisiún na Meán

Government / Regulatory Attendees:

- **CnaM:** Multiple attendees including three Commissioners in person:
  - Niamh Hodnett, Online Safety commissioner
  - John Evans, Digital Services Commissioner
  - Rónán ÓDomhnaill, Media Development Commissioner.
- **European Commission**: Multiple attendees from DG JUST, DG CONNECT, DSA Supervisory Team, EC Representation in Ireland.
- **European Digital Media Observatory (EDMO)**
- **An Garda Síochána**
- **An Coimisiún Toghcháin - The Electoral Commission**
- **Taskforce on Safe Participation in Political Life**

Meta Attendees:

- [HYPERLINK "[redacted]@meta.com" \h], Regulatory Engagement Lead, EICO
- [HYPERLINK "[redacted]@meta.com" \h], Regulatory Risk Management & Insights Lead, Europe

Other Platforms in Attendance:

- TikTok
- Google
- X
- LinkedIn
- Microsoft

Agenda: [HYPERLINK [redacted]

Produced in HJC

CONFIDENTIAL TREATMENT REQUESTED - NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

TL;DR

The roundtable served as an important platform for Meta to reiterate its commitment to election integrity in Ireland, aligning with DSA election guidelines.It also informed stakeholders about specific measures taken to mitigate online bullying and harassment and ensure overall candidate safety. Furthermore, Meta was able to underscore its ongoing collaboration with key election stakeholders, including CnaM, the Electoral Commission, law enforcement, and its third-party fact-checking partner, [HYPERLINK "[redacted]" \h].

- **Candidate Safety:** Candidate safety was the main theme of the presentations and roundtable discussions. CnaM emphasized that threats to candidates and their families are a persistent and serious issue - with increasing threats and harassment both online and offline. Such threats discourage people from participating in the democratic process, undermining democracy itself. CnaM will be monitoring this closely and expects Platforms to take these threats very seriously, remain vigilant and act swiftly.
- **Reporting of Online harassment and abuse**: Many politicians and candidates do not report online abuse. Key reasons for not reporting include complex reporting processes, unsatisfactory responses (e.g., content not removed, no response, or stock replies), desensitization to abuse, lack of clarity on legality, and lack of confidence in law enforcement. Many candidates and politicians felt unsupported by both platforms and authorities.
- **Art 34 and Art 35 DSA obligations**: EC DSA Enforcement Team reminded platforms of their Art 34 and Art 35 DSA obligations which require platforms to conduct risk assessment and put in place appropriate measures in place to mitigate these risks. Platforms need to consider risks to civil discourse and electoral processes and this is a priority focus area for the EC
- **Electoral Guidelines**: CnaM and EC noted that these Guidelines are issued under the DSA. The Guidelines outline measures to be taken before, during and after electoral processes, including recommendations on reporting, engaging with external stakeholders and reinforcing internal processes and publishing information on measures that have been taken for certain elections.
- **Code of Conduct on Misinformation:** EC noted that this is a key instrument in the fight against disinformation. The signatories to the Code have agreed to collaborate in times of elections via the Rapid Response System (RRS) which allows fact checkers to flag sensitive information and content. The RRS will be activated on 29 September to 2 November.
- **Roundtable Discussion:**
    - Platforms were asked how they are preparing for the upcoming Irish Presidential Elections.
    - While no questions were asked about individual cases, the roundtable questions can be linked to recent high-profile cases that have impacted Irish politicians/candidates (defamation - Jim Gavin, frauds/scams/impersonation - Heather Humpheries, threats to harm/kill - Simon Harris).

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

- Platforms were asked questions relating to specific obligations under the Electoral Guidelines and whether or platforms would be adhering to these provisions (i.e. publishing post-election reports).

Detailed Meeting Readout

John Evans, Digital Services Commissioner, CnaM:

- Commissioner Evans thanked everyone for attending and emphasised that candidate safety was the main focus of the Roundtable. He noted that there have been a number of elections across Europe in the past 18 months, not only in Ireland but across Europe as well. Across these elections, the online world continues to be a difficult space for politicians and candidates.
- In terms of candidate safety, and particularly in relation to threats of harm, CnaM will be monitoring this closely and is expecting quick responses from platforms.
- Commissioner Evans referenced the [HYPERLINK "[redacted]" \h] ("**Electoral Guidelines**") and noted these Guidelines provide a high level of detail around expectations.

An Coimisiún Toghcháin - The Irish Electoral Commission - Mary-Clare O'Sullivan

- An Coimisiún Toghcháin provided an update of their voter education and enrolment efforts, including attendance at the Ploughing Championships, University Freshers Weeks and Electric Picnic festival.
- An Coimisiún Toghcháin presented the results of their National Election and Democracy Study which showed a high level of trust in the Irish Electoral System. Voters also have a good understanding of election process knowledge, although some misconceptions remain (like voters needing their polling card to vote, or to vote using a pencil only).
- In terms of attitudes to politicians:
  - 73% think it's unacceptable to record politicians in public
  - 83% think it's unacceptable to protest outside of politicians homes
  - 79% agreed level of aggression is unacceptable
  - 75% agree that their local TD is competent and hardworking.
- A link to An Coimisiún Toghcháin's presentation is [HYPERLINK "[redacted]"].

European Commission, DSA Enforcement Team, Elizabeth Quinn

- EC reminded platforms of their Art 34 and Art 35 DSA obligations which require platforms to conduct risk assessment and put in place appropriate measures in place to mitigate these risks. Platforms need to consider risks to civil discourse and electoral processes and this is a priority focus area for the EC
- EC noted that the Electoral Guidelines outline measures to be taken before, during and after electoral processes, including recommendations on reporting, engaging with

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

external stakeholders and reinforcing internal processes and publishing information on measures that have been taken for certain elections

- The European Board for Digital Services ("**EBDS**") has a working group on elections and these guidelines.
- The EC noted that [HYPERLINK "https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation" \h] is a key instrument in the fight against disinformation. The Code looks at disinformation from different angles while balancing freedom of speech. The signatories to the Code have agreed to collaborate in times of elections via the Rapid Response System (RRS). The RRS allows fact checkers to flag sensitive information and content that they deem to threaten the electoral processes. The RRS will be activated on 29 September to 2 November.

European Commission - DG Just - European Democracy Shield

- The EC is working on a new initiative called the European Democracy Shield which will provide a safeguard to strengthen democracy. This initiative will be adopted this year and will:
    - strengthen cooperation on the detection of disinformation campaigns;
    - strengthen the European Cooperation Network to ensure the safety of political candidates.
    - establish a new democratic resilience centre;
    - support prebunking, debunking, and a network of fact checkers in all languages;
    - include a free and independent media, and citizen participation;

European Commission - DG JUST - Transparency and Targeting of Political Advertising (TTPA)

- EC DG JUST noted that the TPPA enters into force on 10 October and establishes common standards across internal markets for the targeting and delivery of political ads, both online and offline.
- EC DG JUST noted that they are working on implementation of guidance, which aims to support the different actors and DSCs in their oversight functions.
- There will be detailed arrangements for a European registry of political ads and a portal detailing all upcoming elections.
- Member States are currently in the process of appointing national competent authorities for the TPPA. The authorities have formed a network that has met twice this year.

CnaM - Online Safety Commissioner - Niamh Hodnett

- Candidate safety and threats to democracy are top priorities for CnaM.
- Commissioner Hodnett emphasized that threats to politicians and their families are a persistent and serious issue. Such threats discourage people from participating in the democratic process, undermining democracy itself.
- Commissioner Hodnett noted that we all need to step up and do more and platforms are expected to take these threats very seriously and remain vigilant.

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

CnaM - Presentation of CnaM Research - "On the digital campaign trail: election candidates' online experiences in the 2024 elections".

- The study examined the online experiences of candidates in Ireland's 2024 local and general elections. Full results [HYPERLINK "https://www.cnam.ie/app/uploads/2025/08/CnaM_On-the-digital-campaign-trail.pdf" \h].
- **Research Methodology:** The survey only received a 10% response rate—a limitation noted by both CnaM and participants. While at least three responses were received from each political party, candidate fatigue is a real barrier to higher participation in such studies.
- **Social Media as a Campaign Tool**: 95% of candidates used social media for campaigning, highlighting its importance but also the significant challenges it brings, especially regarding online and offline violence. Facebook was the most used platform, followed by Instagram, X/Twitter, TikTok, LinkedIn, and YouTube.
- **Key Findings:**
    - 48% of local and 59% of general election candidates experienced offensive, abusive, or hateful behaviour online.
    - *Threats to kill* or cause serious harm were received by 24% of local and 21% of general election candidates.
    - *Impersonation*: 10% of local and 18% of general election candidates experienced impersonation online.
    - *Targeted Groups:* Women, migrants, and LGBTQ+ candidates faced higher rates and more severe forms of abuse.
    - *Key triggers for abuse:* Immigration (top issue), LGBTQ+ issues, women's issues, housing, and environment.
    - *Chilling effect:* 15% of local and 26% of general election candidates are less likely to run again due to online abuse; 2% (local) and 5% (general) said they would not run again.
- **Reporting of Abuse:**
    - 60% of LE and 70% of GE candidates did not report abusive content.
    - Reasons: complex reporting processes, unhelpful platform responses, desensitization, lack of clarity on legality, and lack of confidence in law enforcement.
    - When reports were made, responses were often unsatisfactory (e.g., content not removed, no response, or stock replies).
    - Many candidates felt unsupported by both platforms and authorities.
- **Recommendations:**
    - *Regulatory Accountability:* Stronger enforcement of existing regulations, meaningful penalties for non-compliance, and multi-stakeholder forums for standardizing reporting and response.
    - *Support for Candidates:* Training on digital safety, clear guidance on reporting, and practical resources.
    - *Public Education:* Awareness campaigns and digital citizenship programs to foster respectful online engagement.
    - *Platform Improvements:* Better reporting mechanisms, faster response times, and more transparent moderation and appeals processes.
    - *Addressing Anonymity:* Consideration of stricter registration or verified-but-private systems to balance accountability and privacy.

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

- *Collaboration*: Ongoing engagement between regulators, law enforcement, political parties, and civil society to address online and offline harms.

An Garda Síochána - Recent Media Reports re Geo-location of cell phones

- The Guards referred to [HYPERLINK "https://about.rte.ie/2025/09/18/rte-prime-time-investigation-uncovers-location-data-of-thousands-of-smartphones-in-ireland-for-sale/" \h] and noted that data location is the key link between online threats and real world harm. While this is an issue for the IDPC, it is one thing to have personalized ads but the selling of location data is a safety risk.
- Guards noted there has been an individual that has turned up outside the Minister for Justice's house 5 times and this is a real threat and link between online and real world.

Roundtable Discussion

Question: In terms of election preparation, how have you prepared for the upcoming Irish Presidential elections and what measures have you put in place?

TikTok:

- Runs a standard election integrity program, including targeted risk assessments in adherence with the Electoral Guidelines.
- Assembles teams of in market subject matter experts (SMEs) in areas like covert influence, policy, and product, both locally and globally.
- Activated an Irish mission control centre to monitor and respond to issues quickly.
- Deploys an Elections Centre 4-5 weeks before the election to provide users with official voting information.
- Promotes media literacy by directing users searching for election content to official sources.

Google:

- Has maintained a stable approach to elections for several years.
- Ensures voters have access to accurate and authoritative information via info panels in Search and YouTube.
- Uses advanced AI and LLMs to detect and counter abuse and misinformation.
- Provides media literacy campaigns (e.g., "Hit Pause") in all EU languages.
- Introduced digital watermarking (SynthID) for AI-generated content.
- Offers enhanced account security for high-risk users (journalists, candidates).
- Operates a threat intelligence team and DDoS protection (Prophet Shield).

Meta:

- Noted that our approach is similar to other platforms
- As part of the election management process, and also in light of the EC Election Guidelines, we have established an Irish elections team. This team was established prior to the announcement of the elections and is the same team that has been in place

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

since the 2024 elections. So the team has good insights in terms of trends and is kept updated with relevant information including tracking elements on and offline.
- We also updated our election risk assessment in advance of the Presidential elections. This risk analysis is across different areas of risks including civil disclosure, ads, threats etc. We map mitigations to these risks.
- We are also committed to providing users with correct information on the elections and have partnered with the Irish election commission and CnaM on this in regards to in app election notifications. Our in app fact checking program and partnership with the journal continues. Political parties also have dedicated teams to assist with any issues.

Microsoft:
- They have most of the same safety measures as other platforms have mentioned.
- Applies a holistic safety approach from product design through to ongoing monitoring.
- Uses both human and automated content moderation.
- Operates a rapid response system, promising action within 24 hours.

LinkedIn:
- As a real ID platform, there is generally more professional and less abusive content.
- Policies are enforced via community standards.
- Prohibits political ads.
- Maintains a dedicated civic issues team and a 24-hour review process for serious threats.

X:
- When it comes to elections, integrity is a key priority for us and authenticity and safety are key areas of focus.
- Standardized process for elections and for assessing potential risks, in order to provide a proportionate response.
- Has a global team that monitors election activity and our safety team remains alert to any attempts to manipulate the platforms white protecting the political parties.
- X's approach to tackling inaccurate and misleading information is "Community Notes" which allows users to collaboratively flag and contextualize misleading information.

Produced to HJC

Question - How are assessments to threats to candidates conducted and managed and what linguistic offerings do you have?
Google:
- *Harassment & Threats*: Google Search and YouTube enforce longstanding policies against harassment and threatening behaviors. YouTube is rarely cited by candidates as a source of negative interactions, but enforcement is in place if issues arise.
- *Linguistic Diversity:* Google employs content moderators across multiple languages, including three with experience moderating in Irish. Teams have access to advanced translation tools, supporting moderation across Europe's diverse languages.

Meta:

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

- *Civic Threats & Hate Speech*: No significant increase in election-specific threats has been observed, but there has been a rise in hostile and hate speech escalations. Meta's team is monitoring closely and ready to respond rapidly.
- *Ad Policies:* Scam ads and deceptive ads using public figures are not allowed, are against our Community Standards and are removed when detected.
- *Language Support:* Irish language content is less than 1% of total, but Meta has Irish language moderators and translation tools. Their fact-checking network spans 29 organizations across the EU, covering 22 languages.

X:

- *Policy Enforcement:* X applies its civic integrity policy and other standards consistently to all users, including candidates.
- *Risk Assessment:* The platform assesses both on- and off-platform risks, considering the scale and severity of potential harms.
- *Candidate Tracking:* X tracks candidates during election events to prevent spam and manipulation, and welcomes lists from authorities to help monitor specific individuals.
- *Reporting Mechanisms:* Robust reporting systems are in place for users to flag problematic content.

CnaM: There is a disconnect between users online and candidates and the measures that have been put in place.

Q from EDMO: we will be participating in the RRS for the first time so we will have the ability to flag content. You have all outlined policies but having a policy and implementing a policy are two different things. There is a gulf between Content that is a threat and content that violates community standards.

TikTok:

- The Rapid Response System (RRS) is designed for civil society organizations to report content that could impact election integrity. TT prioritizes reports of content that clearly violate community guidelines and are harmful. Some reports fall into a grey area (e.g., covert operations) and require more due diligence before action can be taken. TT aims to act quickly on reports.

Question: How are you ensuring the delivery of accurate electoral and candidate information?

Google:

- Seeks to ensure users see a variety of authoritative sources in Search and YouTube.
- As in previous elections, Google's homepage will link directly to the Irish Electoral Commission to provide official, accurate information.

Meta:

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

- Will again run the "election day reminder" initiative which is one of our longest running initiatives.
- Will launch in-app voter information units, and on election day, users will receive notifications linking to official electoral information and encouraging them to share that they have voted.

Irish Electoral Commission:

- Confirmed that in-app notifications from platforms like Meta drive significant increases in traffic to official election resources, demonstrating their effectiveness.
- Asked X about internal contact channels for reporting issues as they did not have a POC.

X: confirmed the POC and will reach out to the electoral commission.

Question - How do you manage incorrect information on candidates, like scams, or incorrect information the election itself?

X:

- Acts quickly on reports from civil society organizations (CSOs) or authorities.
- Takes reports seriously and, rather than removing content, tags it as misleading to inform user

Google:

- Has longstanding processes for addressing election misinformation, such as inaccurate election dates.
- Uses AI content classifiers to detect and escalate misinformation to relevant teams.
- Employs a threat analysis group, a DSA Article 11 point of contact, a rapid response system, and a priority flagger program on YouTube to address urgent or high-risk misinformation.

Microsoft:

- Reviews information to ensure accuracy.
- Removes content that violates policies; if not, the content is "deranked" (its visibility is reduced).

Produced to HJC

Question: Under the EC's Election Guidelines, which are issued under the DSA, there is a recommendation to publish a post election report. Do you intend to publish this and if not, why not?

TT: We started testing our new election app centre where we monitor elections globally and keep a bulletin on what we are doing including content we removed, how many people viewed our resources and spam and covert disruption. This is published in our transparency centre, so this is a version of this type of reporting. In the EU, we also have our transparency reporting as well.

X: this is not a standard we do, sometimes we publish this, but we can't commit to this at this stage. In some cases, if there is a need to do so we will. We publish all figures in transparency reports which are shared with regulators.

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

CnaM: These reports might not be useful to the wider public, but they are useful to stakeholders in the room today.

Meta: We have two levels of reporting. So like TT mentioned, we report via our transparency reports on a biannual basis. We also conduct internal post election reviews where there is an assessment of the mitigation measures deployed including assessment of performance metrics, lessons learned etc. One report is disclosed externally, the transparency reports, and this other one is internal post election review.

Microsoft: We also have our biannual transparency reporting and have internal meetings to share those findings.

Question: What standards do you apply to online abuse of candidates. Do you apply different standards to address online abuse of candidates or are they the same standards as what applies to general members of the public?

Google: we apply the same standards to candidates and the general public. There is a caveat here that in some cases we may allow content to remain online if it is considered newsworthy or educational. So this might make it more likely for content to remain but in principle the guidelines apply to all.

TT: the devil is in the detail but generally speaking there is a line that if it is crossed the content would not be allowed on TT specifically in the cases of harassment and threats of violence, so physical and emotional, that line if crossed, no matter if public or private figure.

Question: In the context of electoral processes, do you have any specific support mechanisms in place for candidates?
TT: we have a specific email address for political parties to contact us. If there is a defamation allegation, this takes longer as it requires a more detailed review and legal process.

Meta: We will proactively put into place mechanisms to protect candidate accounts. We also have a range of tools to empower them during the elections period, which we offer training to candidates for. Specifically, candidates can filter messages with 'Hidden Words'; this tool enables users to filter potentially offensive words, phrases and emojis from your comments and Direct Message requests so they never have to see them again. There is also a 'Limits', which allows people to hide comments and DM requests from people who don't follow them, or who only recently followed them. which prevents people from using certain words in the comments. Our research shows that a lot of negativity against public figures comes from these two groups of people.

Google: we have security tools we encourage candidates to use, routes to contact us,they can use in app reporting tools as this is the fastest way but again we are available to discuss. But we always encourage candidates to use internal tools first.

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Question: There is reporting fatigue and lack of faith that platforms will take action. We are doing a significant amount of work to socialize reporting mechanisms and educate the public but we still have an issue with reporting. We have also seen recent articles on defamation.

X: Relies on automatic and human review so it is removed. In regards to defamation, this is something that can violate our community standards and can be removed. When it is an opinion of an individual but it is not target or inciting harassment or targeting threat to life, that would be left to people to debate and to Community notes.

Meta: they can report to us via normal channels or via solicitors. We will take action if it violates community standards or local laws.

TT: We are the same, they can report to us via normal channels or via solicitors. We will take action if it violates community standards or local laws.

Question: In terms of threats to life, how do you deal with this?
TikTok: Noted that this is a violation of their policies that they will take immediate action. Tikok has a dedicated law enforcement team, and the Guards, political parties and independent candidates have been provided with a direct phone line to call.

Meta: Noted that like TikTok, we also have a LERT team. Meta has policies in place and will work with the Guards to protect politicians but also candidates, voters, election officials, and election-related gatherings. All these scenarios are covered by our monitoring systems and we track content closely.

Google: We have strategic review and operate a follow the sun model and elevations and when they are at an elevated level of alert. These teams also work closely with physical security teams so there is a connection between them.

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 280

DSA PRESIDENTIAL ELECTION ROUNDTABLE. DUBLIN, WEDNESDAY, SEPTEMBER 24, 2025

Attendees: Google, EDMO, Meta, Tiktok, X (online) and CNMA, EC DG connect, An Garda Siochana, Taskforce on Safe Participation in Political Life

Roles: Mainly election regulation, integrity compliance profiles.

Microsoft & LinkedIn: [redacted] (CELA), [redacted] (Bing), [redacted] (LinkedIn), [redacted] (LinkedIn)

NOTES

Coimisiún na Meán': Public attitudes toward elections and trust in democratic processes

- **Pre-bunking & Debunking:** Emphasized as key educational tools to counter disinformation.
- **Campaign Discussion:** Overview of current efforts and strategies.
- **National Election Democracy Study:** Used surveys to gather public opinion with a dedicated election topic. Survey Insights:
    - Trust in Electoral System results: Ballot secrecy: 95% , Legal conduct of elections: 95%, Fairness of officials: 91%, Election management: 91%
    - Election Knowledge Testing: Respondents evaluated on 6 voting-related statements (true/false) including topics on conspiracy theories.
    - Attitudes Toward Politicians: Explored public perceptions and reactions to political recordings.

European Commission Presentation: review DSA a34-35 risk assessment and mitigations

- Focus on civil discourse and electoral processes, and the regulatory frameworks and initiatives to safeguard elections including:
    - **Internal & External Engagement:** Reporting structures, Stakeholder collaboration, Transparency in election-related measures
    - **Working Group on Information Integrity:** Active in monitoring and response.
- COCD focused: Discuss the **rapid Response System** activated during election periods to flag harmful content that will become active next Monday until after the election, supported by EDMO (European Digital Media Observatory).
- The commission is working on a **new Initiative: European Democracy Shield,** covering
    1. Strengthen detection of disinformation campaigns.
    2. Support fact-checking, pre-bunking, debunking across EU languages.
    3. Promote citizen participation and candidate safety.
- **Key Dates** mentioned
    - **Political Ads Regulation to be effective** from **October 10th to** ensure **transparency** and **data protection** (online & offline), informed by: Public consultations, Focus groups, Bilateral meetings
    - **July 9th:** Transiency adopted
    - **April 10th:** Creation of a repository for political ads

Coimisiún na Meán Research Presentation: Candidate experiences and online harms during elections

- **Social Media Use:** 95% of candidates use it for outreach.
- **Research Methods:** Surveys + qualitative interviews (local & general elections)

Key Findings:

- **Negative Online Behavior:** ~50% experienced it
- **Intimidation:** 33% reported threats; 23% reported harmful images
- **Targeting Based on Issues:**
    - Immigration
    - LGBTQ rights
    - Women's reproductive rights
    - Housing
    - Environment
- **Impact of harmful behaviors from bigger to smaller:** Anxiety, safety concerns, increased home security.
- Those who used social media reported following effects on the online behavior made them feel less secure.
- Candidates of SOME and experienced negative online behaviors said they would most likely not run again or if they did they would feel less confident. This was more relevant on women.
- Barriers to reporting negative online experiences: There is skepticism that platforms would take action and what qualifies as legal offense: ~7 out of 10 did NOT report.
- Barriers to Reporting: Skepticism about platform response, Legal ambiguity as a result, ~70% did not report incidents

Themes:

- Support for candidates
- Link between online and offline harm
- Response to violent/intimidatory behavior
- Balancing freedom of expression with civic discourse

Recommendations:

1. **Regulatory:** Address illegal content and threats
2. **Practical Support:** Guidance, education, reporting tools
3. **Partnerships:** Collaborate with Garda Síochána, Oireachtas, local authorities, CSOs, EC

Links to report:

Online experiences negatively affect election candidates, according to our research - Coimisiún na Meán

MICROSOFT CONFIDENTIAL MSFT_JUD_000004176

QUESTIONS

(asked by Coimisiún na Meán Unless stated otherwise)

Q. How do platforms prepare for elections?

Q. Development assessment of threats and linguistic offerings. (There is a gap between this and the feeling of candidates.)

Q. EDMO will be participating in the rapid response system: What content should be flagged? Is it content that violets platform policy, or harms.

Q. How to support electoral candidate inform on accurate process information amplification.

Q. Asks for feedback from platforms. Question for X, is there a specific contact point?

Q. Management of incorrect information on candidate or process.

Q. DSA asks if there is a post-election report.

Q. What has been done for candidate abuse online? Do we have different standards?

Q. Do we have specific support for candidates.

Q. Taskforce on safe participation in political life -- this has to be more platform side.

Q. In terms of defamation, what tools are we dealing with? In Ireland the law goes on top of the community ones.

Q. Have we identified threats to life of candidates?

KEY TAKAWAYS

Broad concern about how digital platforms can proactively safeguard democratic participation, ensure information integrity, protect candidates from harm and examine how online harms translate into offline harm.

Research shows that candidates feel insecure due to the negative online experience behaviors and intimidation with some specific topics on immigration/LGBTQ, Women's issues (e.g. reproductive rights) , and housing. They feel unprotected from platforms that they think do not take actions on this.

Questions raised on platform accountability and preparedness in the context of elections:

- **Platform responsibilities** before, during, and after elections
- **Support mechanisms for candidates**, especially in relation to online abuse, misinformation, and defamation.
- **Consistency and clarity of standards**, including legal frameworks and platform policies.
- **Collaboration with regulatory bodies** and alignment with initiatives like the DSA and EDMO.
- **Focus on Irish context including language.**

MICROSOFT CONFIDENTIAL MSFT_JUD_000004177

Exhibit 281

T: +353 (0)1 644 1200 | info@cnam.ie | www.cnam.ie

Coimisiún na Meán

Risk assessment for Irish Elections

1. Have you conducted a risk assessment for the Irish election? At what point has this been conducted- when the election is officially called or beforehand?
2. What risks were identified?
3. How does the risk profile of the Irish General Election compare to other EU national elections?
4. In relation to conducting the risk assessment and possible mitigation measures is there third-party cooperation in place, in general and for this specific election? For example, cooperation with researchers, CSOs or fact-checkers?

Processes & Procedures

1. What lessons are you taking from the EP elections to national elections?
2. Do you consult with stakeholders on the measures that you will take? If you do so, how does this process work in practice?
3. What is the impact on resources due to other elections occurring at this time?
4. Where is the point of contact for the Irish Elections based?
5. Are there members of the team focusing on Irish elections who speak Irish?
6. What resources are available to the team focusing on Irish Elections?

Candidate Safety

1. What policies and processes do you have in place relating to candidate protection and safety online?
2. If a candidate attack/harassment is coordinated online how would you address it?
3. When candidates report abuse, what is the response time they can expect?
4. What are the policies in place to deal with impersonation of political candidates?

Media Literacy

1. What media literacy initiatives do you have in place?
2. Have you made use of the 2021 Media Literacy Toolbox?

Advertising

1. Are there any proactive measures in place to identify ads with false candidate/party information?
2. If an ad is promoted featuring false candidate/party information how would that be addressed at the content and account level?

Produced to HJC

Coimisiún na Meán
1 Áras Shíol Bhroinn,
Bóthar Shíol Bhroinn,
Baile Átha Cliath 4,
D04 NP20,
Éire.

1 Shelbourne Buildings,
Shelbourne Road,
Dublin 4,
D04 NP20,
Ireland.

Cathaoirleach Feidhmiúcháin: Jeremy Godfrey
Executive Chairperson

Coimisinéirí: Niamh Hodnett | Rónán Ó Domhnaill | John Evans | Aoife MacEvilly
Commissioners

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Coimisiún na Meán

3. What resources are available to review political ads and advertisers when targeting Ireland?

Data Access

1. What avenues do you have in place in relation to transparency and in relation to data access for researchers?

Critical Incidents

1. How do you define a crisis or critical incident?
2. What processes and procedures do you have in place in relation to potential incidents and crises during the electoral period?
3. In the event of violent protests during the electoral period, what are you doing?
4. Do you have 24/7 staff coverage in the event of emergencies? How is this triggered?

Produced to HJC

CONFIDENTIAL TREATMENT REQUESTED -
NOT FOR DISTRIBUTION - MEMBERS & STAFF ONLY

Exhibit 282

	A	B	C	D
1	**Platform**	**Page URL**	**Ad/Content URL**	**Language**
2	**TikTok**	[REDACTED]	[REDACTED]	**Romanian**
3	TikTok	[REDACTED]	[REDACTED]	Romanian
4	TikTok	[REDACTED]	[REDACTED]	Romanian
5	TikTok	[REDACTED]	[REDACTED]	Romanian
6	TikTok	[REDACTED]	[REDACTED]	Romanian
7	TikTok	[REDACTED]	[REDACTED]	Romanian
8	TikTok	[REDACTED]	[REDACTED]	Romanian
9	TikTok	[REDACTED]	[REDACTED]	Romanian
10	TikTok	[REDACTED]	[REDACTED]	Romanian
11	TikTok	[REDACTED]	[REDACTED]	Romanian
12	TikTok	[REDACTED]	[REDACTED]	Romanian
13	TikTok	[REDACTED]	[REDACTED]	Romanian
14	TikTok	[REDACTED]	[REDACTED]	Romanian
15	TikTok	[REDACTED]	[REDACTED]	Romanian
16	TikTok	[REDACTED]	[REDACTED]	Romanian
17	TikTok	[REDACTED]	[REDACTED]	Romanian
18	TikTok	[REDACTED]	[REDACTED]	Romanian
19	TikTok	[REDACTED]	[REDACTED]	Romanian
20	TikTok	[REDACTED]	[REDACTED]	Romanian
21	TikTok	[REDACTED]	[REDACTED]	Romanian
22	TikTok	[REDACTED]	[REDACTED]	Romanian
23	TikTok	[REDACTED]	[REDACTED]	Romanian
24	TikTok	[REDACTED]	[REDACTED]	Romanian
25	TikTok	[REDACTED]	[REDACTED]	Romanian
26	TikTok	[REDACTED]	[REDACTED]	Romanian
27	TikTok	[REDACTED]	[REDACTED]	Romanian
28	TikTok	[REDACTED]	[REDACTED]	Romanian
29	TikTok	[REDACTED]	[REDACTED]	Romanian
30	TikTok	[REDACTED]	[REDACTED]	Romanian
31	TikTok	[REDACTED]	[REDACTED]	Romanian
32	TikTok	[REDACTED]	[REDACTED]	Romanian
33	TikTok	[REDACTED]	[REDACTED]	Romanian
34	TikTok	[REDACTED]	[REDACTED]	Romanian
35	TikTok	[REDACTED]	[REDACTED]	Romanian
36	TikTok	[REDACTED]	[REDACTED]	Romanian
37	TikTok	[REDACTED]	[REDACTED]	Romanian
38	TikTok	[REDACTED]	[REDACTED]	Romanian
39	TikTok	[REDACTED]	[REDACTED]	Romanian
40	TikTok	[REDACTED]	[REDACTED]	Romanian
41	TikTok	[REDACTED]	[REDACTED]	Romanian
42	TikTok	[REDACTED]	[REDACTED]	Romanian
43	TikTok	[REDACTED]	[REDACTED]	Romanian
44	TikTok	[REDACTED]	[REDACTED]	Romanian
45	TikTok	[REDACTED]	[REDACTED]	Romanian
46	TikTok	[REDACTED]	[REDACTED]	Romanian
47	TikTok	[REDACTED]	[REDACTED]	Romanian
48	TikTok	[REDACTED]	[REDACTED]	Romanian
49	TikTok	[REDACTED]	[REDACTED]	Romanian
50	TikTok	[REDACTED]	[REDACTED]	Romanian
51	TikTok	[REDACTED]	[REDACTED]	Romanian
52	TikTok	[REDACTED]	[REDACTED]	Romanian
53	TikTok	[REDACTED]	[REDACTED]	Romanian
54	TikTok	[REDACTED]	[REDACTED]	Romanian
55	TikTok	[REDACTED]	[REDACTED]	Romanian
56	TikTok	[REDACTED]	[REDACTED]	Romanian
57	TikTok	[REDACTED]	[REDACTED]	Romanian
58	TikTok	[REDACTED]	[REDACTED]	Romanian
59	TikTok	[REDACTED]	[REDACTED]	Romanian
60	TikTok	[REDACTED]	[REDACTED]	Romanian

	E
1	**Summary**
2	**This post falsely claims, "Five STS officers have been arrested for vote fraud."**
3	edical assistance, suggesting that meditation is the key solution. The message aligns with a pseudoscientific narrative that may encourage people
4	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
5	In this post, legionism and the Legion of Zelea Codreanu are promoted. The symbols associated with them are prohibited in Romania.
6	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
7	ely claims that Călin Georgescu is already the president and alleges that the protests in Romania are paid events orchestrated as part of a broade
8	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
9	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
10	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
11	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
12	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
13	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
14	post propagates imagery and narratives associated with the military, explicitly linking it to Călin Georgescu, as part of a coordinated propaganda e
15	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
16	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
17	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
18	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
19	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
20	This post claims that AI-generated content is being shared, featuring a fictional conversation between Klaus Iohannis and Călin Georgescu.
21	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
22	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
23	res Călin Georgescu discussing children from Ukraine and their treatment in Romania, while also promoting the narrative that the war in Ukraine
24	Anti-LGBT and Elena Lasconi propaganda
25	Political propaganda for a specific political party outside the campaign
26	Political propaganda for a specific political party outside the campaign
27	Manipulative content related to elections
28	Manipulative content related to elections
29	Manipulative content related to elections
30	Misinformation about elections
31	Misinformation about elections
32	Manipulative content related to elections
33	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
34	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
35	Manipulative content related to elections
36	Manipulative content related to elections
37	Manipulative content related to elections
38	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
39	Misinformation about elections
40	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
41	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
42	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
43	Manipulative content related to elections
44	Manipulative content related to elections
45	Manipulative content related to elections
46	Misinformation about elections
47	Manipulative content related to elections
48	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
49	Manipulative content related to elections
50	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
51	Misinformation about elections
52	Manipulative content related to elections
53	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
54	Manipulative content related to elections
55	Manipulative content related to elections
56	Manipulative content related to elections
57	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
58	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
59	Manipulative content related to elections
60	Misinformation about elections

	F	G
1	**Organisation flagging**	**Reporting date**
2	**BROD - FC**	**01.12.2024**
3	BROD - FC	01.12.2024
4	BROD - FC	01.12.2024
5	BROD - FC	01.12.2024
6	BROD - FC	01.12.2024
7	BROD - FC	01.12.2024
8	BROD - FC	01.12.2024
9	BROD - FC	01.12.2024
10	BROD - FC	01.12.2024
11	BROD - FC	01.12.2024
12	BROD - FC	01.12.2024
13	BROD - FC	01.12.2024
14	BROD - FC	01.12.2024
15	BROD - FC	01.12.2024
16	BROD - FC	01.12.2024
17	BROD - FC	01.12.2024
18	BROD - FC	01.12.2024
19	BROD - FC	01.12.2024
20	BROD - FC	01.12.2024
21	BROD - FC	01.12.2024
22	BROD - FC	01.12.2024
23	BROD - FC	01.12.2024
24	BROD - FC	01.12.2024
25	BROD - FC	01.12.2024
26	BROD - FC	01.12.2024
27	BROD - FC	01/12/2024
28	BROD - FC	01/12/2024
29	BROD - FC	01/12/2024
30	BROD - FC	01/12/2024
31	BROD - FC	01/12/2024
32	BROD - FC	01/12/2024
33	BROD - FC	01/12/2024
34	BROD - FC	01/12/2024
35	BROD - FC	01/12/2024
36	BROD - FC	01/12/2024
37	BROD - FC	01/12/2024
38	BROD - FC	01/12/2024
39	BROD - FC	01/12/2024
40	BROD - FC	01/12/2024
41	BROD - FC	01/12/2024
42	BROD - FC	01/12/2024
43	BROD - FC	01/12/2024
44	BROD - FC	01/12/2024
45	BROD - FC	01/12/2024
46	BROD - FC	01/12/2024
47	BROD - FC	01/12/2024
48	BROD - FC	01/12/2024
49	BROD - FC	01/12/2024
50	BROD - FC	01/12/2024
51	BROD - FC	01/12/2024
52	BROD - FC	01/12/2024
53	BROD - FC	01/12/2024
54	BROD - FC	01/12/2024
55	BROD - FC	01/12/2024
56	BROD - FC	01/12/2024
57	BROD - FC	01/12/2024
58	BROD - FC	01/12/2024
59	BROD - FC	01/12/2024
60	BROD - FC	01/12/2024

	A	B	C	D
61	TikTok	[REDACTED]	[REDACTED]	Romanian
62	TikTok	[REDACTED]	[REDACTED]	Romanian
63	TikTok	[REDACTED]	[REDACTED]	Romanian
64	TikTok	[REDACTED]	[REDACTED]	Romanian
65	TikTok	[REDACTED]	[REDACTED]	Romanian
66	TikTok	[REDACTED]	[REDACTED]	Romanian
67	TikTok	[REDACTED]	[REDACTED]	Romanian
68	TikTok	[REDACTED]	[REDACTED]	Romanian
69	TikTok	[REDACTED]	[REDACTED]	Romanian
70	TikTok	[REDACTED]	[REDACTED]	Romanian
71	TikTok	[REDACTED]	[REDACTED]	Romanian
72	TikTok	[REDACTED]	[REDACTED]	Romanian
73	TikTok	[REDACTED]	[REDACTED]	Romanian
74	TikTok	[REDACTED]	[REDACTED]	Romanian
75	TikTok	[REDACTED]	[REDACTED]	Romanian
76	TikTok	[REDACTED]	[REDACTED]	Romanian
77	TikTok	[REDACTED]	[REDACTED]	Romanian
78	TikTok	[REDACTED]	[REDACTED]	Romanian
79	TikTok	[REDACTED]	[REDACTED]	Romanian
80	TikTok	[REDACTED]	[REDACTED]	Romanian
81	TikTok	[REDACTED]	[REDACTED]	Romanian
82	TikTok	[REDACTED]	[REDACTED]	Romanian
83	TikTok	[REDACTED]	[REDACTED]	Romanian
84	TikTok	[REDACTED]	[REDACTED]	Romanian
85	TikTok	[REDACTED]	[REDACTED]	Romanian
86	TikTok	[REDACTED]	[REDACTED]	Romanian
87	TikTok	[REDACTED]	[REDACTED]	Romanian
88	TikTok	[REDACTED]	[REDACTED]	Romanian
89	TikTok	[REDACTED]	[REDACTED]	Romanian
90	TikTok	[REDACTED]	[REDACTED]	Romanian
91	TikTok	[REDACTED]	[REDACTED]	Romanian
92	TikTok	[REDACTED]	[REDACTED]	Romanian
93	TikTok	[REDACTED]	[REDACTED]	Romanian
94	TikTok	[REDACTED]	[REDACTED]	Romanian
95	TikTok	[REDACTED]	[REDACTED]	Romanian
96	TikTok	[REDACTED]	[REDACTED]	Romanian
97	TikTok	[REDACTED]	[REDACTED]	Romanian
98	TikTok	[REDACTED]	[REDACTED]	Romanian
99	TikTok	[REDACTED]	[REDACTED]	Romanian
100	TikTok	[REDACTED]	[REDACTED]	Romanian
101	TikTok	[REDACTED]	[REDACTED]	Romanian
102	TikTok	[REDACTED]	[REDACTED]	Romanian
103	TikTok	[REDACTED]	[REDACTED]	Romanian
104	TikTok	[REDACTED]	[REDACTED]	Romanian
105	TikTok	[REDACTED]	[REDACTED]	Romanian
106	TikTok	[REDACTED]	[REDACTED]	Romanian
107	TikTok	[REDACTED]	[REDACTED]	Romanian
108	TikTok	[REDACTED]	[REDACTED]	Romanian
109	TikTok	[REDACTED]	[REDACTED]	Romanian
110	TikTok	[REDACTED]	[REDACTED]	Romanian
111	TikTok	[REDACTED]	[REDACTED]	Romanian
112	TikTok	[REDACTED]	[REDACTED]	Romanian
113	TikTok	[REDACTED]	[REDACTED]	Romanian
114	TikTok	[REDACTED]	[REDACTED]	Romanian
115	TikTok	[REDACTED]	[REDACTED]	Romanian
116	TikTok	[REDACTED]	[REDACTED]	Romanian
117	TikTok	[REDACTED]	[REDACTED]	Romanian
118	TikTok	[REDACTED]	[REDACTED]	Romanian
119	TikTok	[REDACTED]	[REDACTED]	Romanian
120	TikTok	[REDACTED]	[REDACTED]	Romanian

	E
61	Manipulative content related to elections
62	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
63	Manipulative content related to elections
64	Manipulative content related to elections
65	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
66	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
67	Manipulative content related to elections
68	Manipulative content related to elections
69	Manipulative content related to elections
70	Manipulative content related to elections
71	Misinformation about elections
72	Manipulative content related to elections
73	Manipulative content related to elections
74	Misinformation about elections
75	Manipulative content related to elections
76	Manipulative content related to elections
77	Misinformation about elections
78	Manipulative content related to elections
79	Manipulative content related to elections
80	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
81	Manipulative content related to elections
82	Manipulative content related to elections
83	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
84	Manipulative content related to elections
85	Misinformation about elections
86	Manipulative content related to elections
87	Manipulative content related to elections
88	Manipulative content related to elections
89	Manipulative content related to elections
90	Manipulative content related to elections
91	Manipulative content related to elections
92	Manipulative content related to elections
93	Manipulative content related to elections
94	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
95	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
96	Manipulative content related to elections
97	Manipulative content related to elections
98	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
99	Manipulative content related to elections
100	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
101	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
102	Manipulative content related to elections
103	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
104	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
105	Misinformation about elections
106	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
107	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
108	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
109	Manipulative content related to elections
110	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
111	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
112	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
113	Manipulative content related to elections
114	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
115	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
116	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
117	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
118	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
119	Manipulative content related to elections
120	Manipulative content related to elections

	F	G
61	BROD - FC	01/12/2024
62	BROD - FC	01/12/2024
63	BROD - FC	01/12/2024
64	BROD - FC	01/12/2024
65	BROD - FC	01/12/2024
66	BROD - FC	01/12/2024
67	BROD - FC	01/12/2024
68	BROD - FC	01/12/2024
69	BROD - FC	01/12/2024
70	BROD - FC	01/12/2024
71	BROD - FC	01/12/2024
72	BROD - FC	01/12/2024
73	BROD - FC	01/12/2024
74	BROD - FC	01/12/2024
75	BROD - FC	01/12/2024
76	BROD - FC	01/12/2024
77	BROD - FC	01/12/2024
78	BROD - FC	01/12/2024
79	BROD - FC	01/12/2024
80	BROD - FC	01/12/2024
81	BROD - FC	01/12/2024
82	BROD - FC	01/12/2024
83	BROD - FC	01/12/2024
84	BROD - FC	01/12/2024
85	BROD - FC	01/12/2024
86	BROD - FC	01/12/2024
87	BROD - FC	01/12/2024
88	BROD - FC	01/12/2024
89	BROD - FC	01/12/2024
90	BROD - FC	01/12/2024
91	BROD - FC	01/12/2024
92	BROD - FC	01/12/2024
93	BROD - FC	01/12/2024
94	BROD - FC	01/12/2024
95	BROD - FC	01/12/2024
96	BROD - FC	01/12/2024
97	BROD - FC	01/12/2024
98	BROD - FC	01/12/2024
99	BROD - FC	01/12/2024
100	BROD - FC	01/12/2024
101	BROD - FC	01/12/2024
102	BROD - FC	01/12/2024
103	BROD - FC	01/12/2024
104	BROD - FC	01/12/2024
105	BROD - FC	01/12/2024
106	BROD - FC	01/12/2024
107	BROD - FC	01/12/2024
108	BROD - FC	01/12/2024
109	BROD - FC	01/12/2024
110	BROD - FC	01/12/2024
111	BROD - FC	01/12/2024
112	BROD - FC	01/12/2024
113	BROD - FC	01/12/2024
114	BROD - FC	01/12/2024
115	BROD - FC	01/12/2024
116	BROD - FC	01/12/2024
117	BROD - FC	01/12/2024
118	BROD - FC	01/12/2024
119	BROD - FC	01/12/2024
120	BROD - FC	01/12/2024

	A	B	C	D
121	TikTok	[REDACTED]	[REDACTED]	Romanian
122	TikTok	[REDACTED]	[REDACTED]	Romanian
123	TikTok	[REDACTED]	[REDACTED]	Romanian
124	TikTok	[REDACTED]	[REDACTED]	Romanian
125	TikTok	[REDACTED]	[REDACTED]	Romanian
126	TikTok	[REDACTED]	[REDACTED]	Romanian
127	TikTok	[REDACTED]	[REDACTED]	Romanian
128	TikTok	[REDACTED]	[REDACTED]	Romanian
129	TikTok	[REDACTED]	[REDACTED]	Romanian
130	TikTok	[REDACTED]	[REDACTED]	Romanian
131	TikTok	[REDACTED]	[REDACTED]	Romanian
132	TikTok	[REDACTED]	[REDACTED]	Romanian
133	TikTok	[REDACTED]	[REDACTED]	Romanian
134	TikTok	[REDACTED]	[REDACTED]	Romanian
135	TikTok	[REDACTED]	[REDACTED]	Romanian
136	TikTok	[REDACTED]	[REDACTED]	Romanian
137	TikTok	[REDACTED]	[REDACTED]	Romanian
138	TikTok	[REDACTED]	[REDACTED]	Romanian
139	TikTok	[REDACTED]	[REDACTED]	Romanian
140	TikTok	[REDACTED]	[REDACTED]	Romanian
141	TikTok	[REDACTED]	[REDACTED]	Romanian
142	TikTok	[REDACTED]	[REDACTED]	Romanian
143	TikTok	[REDACTED]	[REDACTED]	Romanian
144	TikTok	[REDACTED]	[REDACTED]	Romanian
145	TikTok	[REDACTED]	[REDACTED]	Romanian
146	TikTok	[REDACTED]	[REDACTED]	Romanian
147	TikTok	[REDACTED]	[REDACTED]	Romanian
148	TikTok	[REDACTED]	[REDACTED]	Romanian
149	TikTok	[REDACTED]	[REDACTED]	Romanian
150	TikTok	[REDACTED]	[REDACTED]	Romanian
151	TikTok	[REDACTED]	[REDACTED]	Romanian
152	TikTok	[REDACTED]	[REDACTED]	Romanian
153	TikTok	[REDACTED]	[REDACTED]	Romanian
154	TikTok	[REDACTED]	[REDACTED]	Romanian
155	TikTok	[REDACTED]	[REDACTED]	Romanian
156	TikTok	[REDACTED]	[REDACTED]	Romanian
157	TikTok	[REDACTED]	[REDACTED]	Romanian
158	TikTok	[REDACTED]	[REDACTED]	Romanian
159	TikTok	[REDACTED]	[REDACTED]	Romanian
160	TikTok	[REDACTED]	[REDACTED]	Romanian
161	TikTok	[REDACTED]	[REDACTED]	Romanian
162	TikTok	[REDACTED]	[REDACTED]	Romanian
163	TikTok	[REDACTED]	[REDACTED]	Romanian
164	TikTok	[REDACTED]	[REDACTED]	Romanian
165	TikTok	[REDACTED]	[REDACTED]	Romanian
166	TikTok	[REDACTED]	[REDACTED]	Romanian
167	TikTok	[REDACTED]	[REDACTED]	Romanian
168	TikTok	[REDACTED]	[REDACTED]	Romanian
169	TikTok	[REDACTED]	[REDACTED]	Romanian
170	TikTok	[REDACTED]	[REDACTED]	Romanian
171	TikTok	[REDACTED]	[REDACTED]	Romanian
172	TikTok	[REDACTED]	[REDACTED]	Romanian
173	TikTok	[REDACTED]	[REDACTED]	Romanian
174	TikTok	[REDACTED]	[REDACTED]	Romanian
175	TikTok	[REDACTED]	[REDACTED]	Romanian
176	TikTok	[REDACTED]	[REDACTED]	Romanian
177	TikTok	[REDACTED]	[REDACTED]	Romanian
178	TikTok	[REDACTED]	[REDACTED]	Romanian
179	TikTok	[REDACTED]	[REDACTED]	Romanian
180	TikTok	[REDACTED]	[REDACTED]	Romanian

	E
121	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
122	Manipulative content related to elections
123	Manipulative content related to elections
124	Manipulative content related to elections
125	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
126	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
127	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
128	Manipulative content related to elections
129	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
130	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
131	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
132	Manipulative content related to elections
133	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
134	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
135	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
136	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
137	Manipulative content related to elections
138	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
139	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
140	Manipulative content related to elections
141	Manipulative content related to elections
142	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
143	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
144	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
145	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
146	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
147	Manipulative content related to elections
148	Manipulative content related to elections
149	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
150	Manipulative content related to elections
151	Manipulative content related to elections
152	Manipulative content related to elections
153	Manipulative content related to elections
154	Manipulative content related to elections
155	Manipulative content related to elections
156	Manipulative content related to elections
157	Misinformation about elections
158	Manipulative content related to elections
159	Manipulative content related to elections
160	Manipulative content related to elections
161	Manipulative content related to elections
162	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
163	Misinformation about elections
164	Manipulative content related to elections
165	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
166	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
167	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
168	Manipulative content related to elections
169	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
170	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
171	Manipulative content related to elections
172	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
173	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
174	Misinformation about elections
175	Manipulative content related to elections
176	Manipulative content related to elections
177	Manipulative content related to elections
178	Manipulative content related to elections
179	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
180	Manipulative content related to elections

	F	G
121	BROD - FC	01/12/2024
122	BROD - FC	01/12/2024
123	BROD - FC	01/12/2024
124	BROD - FC	01/12/2024
125	BROD - FC	01/12/2024
126	BROD - FC	01/12/2024
127	BROD - FC	01/12/2024
128	BROD - FC	01/12/2024
129	BROD - FC	01/12/2024
130	BROD - FC	01/12/2024
131	BROD - FC	01/12/2024
132	BROD - FC	01/12/2024
133	BROD - FC	01/12/2024
134	BROD - FC	01/12/2024
135	BROD - FC	01/12/2024
136	BROD - FC	01/12/2024
137	BROD - FC	01/12/2024
138	BROD - FC	01/12/2024
139	BROD - FC	01/12/2024
140	BROD - FC	01/12/2024
141	BROD - FC	01/12/2024
142	BROD - FC	01/12/2024
143	BROD - FC	01/12/2024
144	BROD - FC	01/12/2024
145	BROD - FC	01/12/2024
146	BROD - FC	01/12/2024
147	BROD - FC	01/12/2024
148	BROD - FC	01/12/2024
149	BROD - FC	01/12/2024
150	BROD - FC	01/12/2024
151	BROD - FC	01/12/2024
152	BROD - FC	01/12/2024
153	BROD - FC	01/12/2024
154	BROD - FC	01/12/2024
155	BROD - FC	01/12/2024
156	BROD - FC	01/12/2024
157	BROD - FC	01/12/2024
158	BROD - FC	01/12/2024
159	BROD - FC	01/12/2024
160	BROD - FC	01/12/2024
161	BROD - FC	01/12/2024
162	BROD - FC	01/12/2024
163	BROD - FC	01/12/2024
164	BROD - FC	01/12/2024
165	BROD - FC	01/12/2024
166	BROD - FC	01/12/2024
167	BROD - FC	01/12/2024
168	BROD - FC	01/12/2024
169	BROD - FC	01/12/2024
170	BROD - FC	01/12/2024
171	BROD - FC	01/12/2024
172	BROD - FC	01/12/2024
173	BROD - FC	01/12/2024
174	BROD - FC	01/12/2024
175	BROD - FC	01/12/2024
176	BROD - FC	01/12/2024
177	BROD - FC	01/12/2024
178	BROD - FC	01/12/2024
179	BROD - FC	01/12/2024
180	BROD - FC	01/12/2024

	A	B	C	D
181	TikTok	[REDACTED]	[REDACTED]	Romanian
182	TikTok	[REDACTED]	[REDACTED]	Romanian
183	TikTok	[REDACTED]	[REDACTED]	Romanian
184	TikTok	[REDACTED]	[REDACTED]	Romanian
185	TikTok	[REDACTED]	[REDACTED]	Romanian
186	TikTok	[REDACTED]	[REDACTED]	Romanian
187	TikTok	[REDACTED]	[REDACTED]	Romanian
188	TikTok	[REDACTED]	[REDACTED]	Romanian
189	TikTok	[REDACTED]	[REDACTED]	Romanian
190	TikTok	[REDACTED]	[REDACTED]	Romanian
191	TikTok	[REDACTED]	[REDACTED]	Romanian
192	TikTok	[REDACTED]	[REDACTED]	Romanian
193	TikTok	[REDACTED]	[REDACTED]	Romanian
194	TikTok	[REDACTED]	[REDACTED]	Romanian
195	TikTok	[REDACTED]	[REDACTED]	Romanian
196	TikTok	[REDACTED]	[REDACTED]	Romanian
197	TikTok	[REDACTED]	[REDACTED]	Romanian
198	TikTok	[REDACTED]	[REDACTED]	Romanian
199	TikTok	[REDACTED]	[REDACTED]	Romanian
200	TikTok	[REDACTED]	[REDACTED]	Romanian
201	TikTok	[REDACTED]	[REDACTED]	Romanian
202	TikTok	[REDACTED]	[REDACTED]	Romanian
203	TikTok	[REDACTED]	[REDACTED]	Romanian
204	TikTok	[REDACTED]	[REDACTED]	Romanian
205	TikTok	[REDACTED]	[REDACTED]	Romanian
206	TikTok	[REDACTED]	[REDACTED]	Romanian
207	TikTok	[REDACTED]	[REDACTED]	Romanian
208	TikTok	[REDACTED]	[REDACTED]	Romanian
209	TikTok	[REDACTED]	[REDACTED]	Romanian
210	TikTok	[REDACTED]	[REDACTED]	Romanian
211	TikTok	[REDACTED]	[REDACTED]	Romanian
212	TikTok	[REDACTED]	[REDACTED]	Romanian
213	TikTok	[REDACTED]	[REDACTED]	Romanian
214	TikTok	[REDACTED]	[REDACTED]	Romanian
215	TikTok	[REDACTED]	[REDACTED]	Romanian
216	TikTok	[REDACTED]	[REDACTED]	Romanian
217	TikTok	[REDACTED]	[REDACTED]	Romanian
218	TikTok	[REDACTED]	[REDACTED]	Romanian
219	TikTok	[REDACTED]	[REDACTED]	Romanian
220	TikTok	[REDACTED]	[REDACTED]	Romanian
221	TikTok	[REDACTED]	[REDACTED]	Romanian
222	TikTok	[REDACTED]	[REDACTED]	Romanian
223	TikTok	[REDACTED]	[REDACTED]	Romanian
224	TikTok	[REDACTED]	[REDACTED]	Romanian
225	TikTok	[REDACTED]	[REDACTED]	Romanian
226	TikTok	[REDACTED]	[REDACTED]	Romanian
227	TikTok	[REDACTED]	[REDACTED]	Romanian
228	TikTok	[REDACTED]	[REDACTED]	Romanian
229	TikTok	[REDACTED]	[REDACTED]	Romanian
230	TikTok	[REDACTED]	[REDACTED]	Romanian
231	TikTok	[REDACTED]	[REDACTED]	Romanian
232	TikTok	[REDACTED]	[REDACTED]	Romanian
233	TikTok	[REDACTED]	[REDACTED]	Romanian
234	TikTok	[REDACTED]	[REDACTED]	Romanian
235	TikTok	[REDACTED]	[REDACTED]	Romanian
236	TikTok	[REDACTED]	[REDACTED]	Romanian
237	TikTok	[REDACTED]	[REDACTED]	Romanian
238	TikTok	[REDACTED]	[REDACTED]	Romanian
239	TikTok	[REDACTED]	[REDACTED]	Romanian
240	TikTok	[REDACTED]	[REDACTED]	Romanian

	E
181	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
182	Manipulative content related to elections
183	Manipulative content related to elections
184	Manipulative content related to elections
185	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
186	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
187	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
188	Manipulative content related to elections
189	Manipulative content related to elections
190	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
191	Manipulative content related to elections
192	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
193	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
194	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
195	Manipulative content related to elections
196	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
197	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
198	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
199	Manipulative content related to elections
200	Manipulative content related to elections
201	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
202	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
203	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
204	Manipulative content related to elections
205	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
206	Manipulative content related to elections
207	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
208	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
209	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
210	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
211	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
212	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
213	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
214	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
215	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
216	Manipulative content related to elections
217	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
218	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
219	Manipulative content related to elections
220	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
221	Manipulative content related to elections
222	Manipulative content related to elections
223	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
224	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
225	Manipulative content related to elections
226	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
227	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
228	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
229	Manipulative content related to elections
230	Manipulative content related to elections
231	Manipulative content related to elections
232	Manipulative content related to elections
233	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
234	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
235	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)
236	Misinformation about elections
237	Misinformation about elections
238	Misinformation about elections
239	Misinformation about elections
240	Misinformation about elections

	F	G
181	BROD - FC	01/12/2024
182	BROD - FC	01/12/2024
183	BROD - FC	01/12/2024
184	BROD - FC	01/12/2024
185	BROD - FC	01/12/2024
186	BROD - FC	01/12/2024
187	BROD - FC	01/12/2024
188	BROD - FC	01/12/2024
189	BROD - FC	01/12/2024
190	BROD - FC	01/12/2024
191	BROD - FC	01/12/2024
192	BROD - FC	01/12/2024
193	BROD - FC	01/12/2024
194	BROD - FC	01/12/2024
195	BROD - FC	01/12/2024
196	BROD - FC	01/12/2024
197	BROD - FC	01/12/2024
198	BROD - FC	01/12/2024
199	BROD - FC	01/12/2024
200	BROD - FC	01/12/2024
201	BROD - FC	01/12/2024
202	BROD - FC	01/12/2024
203	BROD - FC	01/12/2024
204	BROD - FC	01/12/2024
205	BROD - FC	01/12/2024
206	BROD - FC	01/12/2024
207	BROD - FC	01/12/2024
208	BROD - FC	01/12/2024
209	BROD - FC	01/12/2024
210	BROD - FC	01/12/2024
211	BROD - FC	01/12/2024
212	BROD - FC	01/12/2024
213	BROD - FC	01/12/2024
214	BROD - FC	01/12/2024
215	BROD - FC	01/12/2024
216	BROD - FC	01/12/2024
217	BROD - FC	01/12/2024
218	BROD - FC	01/12/2024
219	BROD - FC	01/12/2024
220	BROD - FC	01/12/2024
221	BROD - FC	01/12/2024
222	BROD - FC	01/12/2024
223	BROD - FC	01/12/2024
224	BROD - FC	01/12/2024
225	BROD - FC	01/12/2024
226	BROD - FC	01/12/2024
227	BROD - FC	01/12/2024
228	BROD - FC	01/12/2024
229	BROD - FC	01/12/2024
230	BROD - FC	01/12/2024
231	BROD - FC	01/12/2024
232	BROD - FC	01/12/2024
233	BROD - FC	01/12/2024
234	BROD - FC	01/12/2024
235	BROD - FC	01/12/2024
236	BROD - FC	01/12/2024
237	BROD - FC	01/12/2024
238	BROD - FC	01/12/2024
239	BROD - FC	01/12/2024
240	BROD - FC	01/12/2024

	A	B	C	D
241	TikTok	[REDACTED]	[REDACTED]	Romanian
242	TikTok	[REDACTED]	[REDACTED]	Romanian

	E
241	Political propaganda for Calin Georgescu
242	Another post includes a video discussing Emoto Masaru and the concept of water memory, accompanied by the message below the video: "Votă

	F	G
241	BROD - FC	01/12/2024
242	BROD - FC	01/12/2024

Exhibit 283

	A	B	C	D
1	**Platform**	**Page URL**	**Ad/Content URL**	**Language**
2	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
3	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
4	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
5	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
6	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
7	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
8	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
9	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
10	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
11	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
12	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
13	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
14	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
15	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
16	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
17	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
18	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
19	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
20	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
21	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
22	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
23	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
24	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
25	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
26	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
27	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
28	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
29	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
30	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
31	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
32	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
33	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
34	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
35	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
36	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
37	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
38	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
39	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
40	TikTok	[REDACTED]	[REDACTED]	ROMANIAN
41	TikTok	[REDACTED]	[REDACTED]	ROMANIAN

	E
1	**Summary**
2	Deepfake of candidate Călin Georgescu claiming to support the opposing party.
3	This post promotes conspiracy theories that one candidate previously endorsed during their political campaign for the presidency.
4	This post spreads disinformation about the LGBTQ community and religion in schools.
5	The post claims that the state is preparing something concerning tomorrow's vote because the author saw police and military personnel on the st
6	This account belongs to someone promoting pro-Russia, anti-EU, and pro-Georgescu rhetoric. Their accounts are frequently shut down, leading
7	Another very popular account with 498.4K followers. In the video, it claims that "we export a kilogram of generic wheat for 80 bani, buy frozen do
8	It's a post where Georgescu states that "Diaspora is a completely incorrect term" used for Romanians living abroad. The account features severa
9	This post promotes Călin Georgescu's narratives and contains multiple instances of disinformation presented in the video.
10	The post is calling for a revolution if the election is annulled.
11	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
12	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
13	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
14	All the posts on this account are disinformation, as they align Călin Georgescu's ideology with the ideology of SENS (political party). They promot
15	An "informative" video attempting to explain what Călin Georgescu meant when he spoke about the "breaking of the divine thread through C-sect
16	Propaganda for Calin Georgescu
17	A viral AI-generated video (167.4k likes) claims that "Călin Georgescu's plan will completely transform the country in just 4 years." It makes outra
18	Propaganda for Calin Georgescu
19	Propaganda for Calin Georgescu
20	Propaganda for Calin Georgescu
21	Propaganda for Calin Georgescu
22	Propaganda for Calin Georgescu
23	Propaganda for Calin Georgescu
24	Propaganda for Calin Georgescu
25	Propaganda for Calin Georgescu
26	Propaganda for Calin Georgescu
27	A post featuring a prediction attributed to Baba Vanga about a messianic leader who will transform Romania into an exemplary nation for the worl
28	A slide explaining CG's statements, with slide 6 claiming that USR "locked us in our homes, put the dead in bags..." This post has garnered 27.8
29	Propaganda for Calin Georgescu. He claims he would resolve the Schengen issue within 24 hours (plus additional commentary about highways a
30	An account with 67.4k followers features an experiment connecting water pH measurements to CG's theory about water's "information," claiming
31	This post claims: "Robert Kennedy Jr. is coming to Romania to support Călin Georgescu. Trump also supports CG."
32	This post promotes a conspiracy theory about the elections and connects it to the ongoing protests in Romania.
33	Propaganda for Political Party POT - Partidul Oamenilor Tineri - outside the campaign
34	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
35	Propaganda for Political Party POT - Partidul Oamenilor Tineri - outside the campaign
36	Propaganda for Political Party POT - Partidul Oamenilor Tineri - outside the campaign
37	The post claims that nanoparticle technology in the food industry is used to control the human body and behavior.
38	Propaganda for Political Party POT - Partidul Oamenilor Tineri - outside the campaign
39	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
40	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
41	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.

	F	G
1	**Organisation flagging**	**Reporting date**
2	BROD - FC	30.11.2024
3	BROD - FC	30.11.2024
4	BROD - FC	30.11.2024
5	BROD - FC	30.11.2024
6	BROD - FC	30.11.2024
7	BROD - FC	30.11.2024
8	BROD - FC	30.11.2024
9	BROD - FC	30.11.2024
10	BROD - FC	30.11.2024
11	BROD - FC	30.11.2024
12	BROD - FC	30.11.2024
13	BROD - FC	30.11.2024
14	BROD - FC	30.11.2024
15	BROD - FC	30.11.2024
16	BROD - FC	30.11.2024
17	BROD - FC	30.11.2024
18	BROD - FC	30.11.2024
19	BROD - FC	30.11.2024
20	BROD - FC	30.11.2024
21	BROD - FC	30.11.2024
22	BROD - FC	30.11.2024
23	BROD - FC	30.11.2024
24	BROD - FC	30.11.2024
25	BROD - FC	30.11.2024
26	BROD - FC	30.11.2024
27	BROD - FC	30.11.2024
28	BROD - FC	30.11.2024
29	BROD - FC	30.11.2024
30	BROD - FC	30.11.2024
31	BROD - FC	30.11.2024
32	BROD - FC	30.11.2024
33	BROD - FC	30.11.2024
34	BROD - FC	30.11.2024
35	BROD - FC	30.11.2024
36	BROD - FC	30.11.2024
37	BROD - FC	30.11.2024
38	BROD - FC	30.11.2024
39	BROD - FC	30.11.2024
40	BROD - FC	30.11.2024
41	BROD - FC	30.11.2024

Exhibit 284

	A	B	C	D
1	**Platform**	**Page URL**	**Ad/Content URL**	**Language**
2	TikTok	[REDACTED]	[REDACTED]	Romanian
3	TikTok	[REDACTED]	[REDACTED]	Romanian
4	TikTok	[REDACTED]	[REDACTED]	Romanian
5	TikTok	[REDACTED]	[REDACTED]	Romanian
6	TikTok	[REDACTED]	[REDACTED]	Romanian
7	TikTok	[REDACTED]	[REDACTED]	Romanian
8	TikTok	[REDACTED]	[REDACTED]	Romanian
9	TikTok	[REDACTED]	[REDACTED]	Romanian
10	TikTok	[REDACTED]	[REDACTED]	Romanian
11	TikTok	[REDACTED]	[REDACTED]	Romanian
12	TikTok	[REDACTED]	[REDACTED]	Romanian
13	TikTok	[REDACTED]	[REDACTED]	Romanian
14	TikTok	[REDACTED]	[REDACTED]	Romanian
15	TikTok	[REDACTED]	[REDACTED]	Romanian
16	TikTok	[REDACTED]	[REDACTED]	Romanian
17	TikTok	[REDACTED]	[REDACTED]	Romanian
18	TikTok	[REDACTED]	[REDACTED]	Romanian
19	TikTok	[REDACTED]	[REDACTED]	Romanian
20	TikTok	[REDACTED]	[REDACTED]	Romanian
21	TikTok	[REDACTED]	[REDACTED]	Romanian
22	TikTok	[REDACTED]	[REDACTED]	Romanian
23	TikTok	[REDACTED]	[REDACTED]	Romanian
24	TikTok	[REDACTED]	[REDACTED]	Romanian
25	TikTok	[REDACTED]	[REDACTED]	Romanian
26	TikTok	[REDACTED]	[REDACTED]	Romanian
27	TikTok	[REDACTED]	[REDACTED]	Romanian
28	TikTok	[REDACTED]	[REDACTED]	Romanian
29	TikTok	[REDACTED]	[REDACTED]	Romanian
30	TikTok	[REDACTED]	[REDACTED]	Romanian
31	TikTok	[REDACTED]	[REDACTED]	Romanian
32	TikTok	[REDACTED]	[REDACTED]	Romanian

	E
1	**Summary**
2	In this post, legionism and the Legion of Zelea Codreanu are promoted. The symbols associated with them are prohibited in Romania.
3	In this post, legionism and the Legion of Zelea Codreanu are promoted. The symbols associated with them are prohibited in Romania.
4	In this post, legionism and the Legion of Zelea Codreanu are promoted. The symbols associated with them are prohibited in Romania.
5	In this post, legionism and the Legion of Zelea Codreanu are promoted. The symbols associated with them are prohibited in Romania.
6	In this post, legionism and the Legion of Zelea Codreanu are promoted. The symbols associated with them are prohibited in Romania.
7	In this post, legionism and the Legion of Zelea Codreanu are promoted. The symbols associated with them are prohibited in Romania.
8	This post discusses the development of information related to water, referencing Masaru Emoto's study. It claims that bottled water becomes ha
9	This post discusses the development of information related to water, referencing Masaru Emoto's study. It claims that bottled water becomes ha
10	This post claims that television acknowledges the existence of nanochips in water, referring to microplastics. It aligns with the narrative promoted
11	This post suggests that in Sweden, newborns are not separated from the umbilical cord for tens of minutes or even hours. It aligns with the narra
12	This post highlights Călin Georgescu's claim during a podcast with Micutzu that he could resolve the Schengen issue within 24 hours. He argues
13	This post on this account are disinformation, as they align Călin Georgescu's ideology with the ideology of SENS (political party). They promote v
14	This post on this account are disinformation, as they align Călin Georgescu's ideology with the ideology of SENS (political party). They promote v
15	This post on this account are disinformation, as they align Călin Georgescu's ideology with the ideology of SENS (political party). They promote v
16	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
17	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
18	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
19	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
20	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
21	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
22	This post constitutes disinformation targeting a presidential candidate, falsely claiming that she will actively promote an LGBT agenda if elected.
23	This post constitutes disinformation targeting a presidential candidate, falsely claiming that she will actively promote an LGBT agenda if elected.
24	This post constitutes disinformation targeting a presidential candidate, falsely claiming that she will actively promote an LGBT agenda if elected.
25	Propaganda for Calin Georgescu
26	Propaganda for Calin Georgescu
27	Propaganda for Calin Georgescu
28	Propaganda for Calin Georgescu
29	This post on this account are disinformation, as they align Călin Georgescu's ideology with the ideology of SENS (political party). They promote v
30	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
31	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.
32	The post promotes propaganda for a specific political party and claims that the secrecy of the vote is not being respected.

	F	G
1	**Organisation flagging**	**Reporting date**
2	BROD - FC	01.12.2024
3	BROD - FC	01.12.2024
4	BROD - FC	01.12.2024
5	BROD - FC	01.12.2024
6	BROD - FC	01.12.2024
7	BROD - FC	01.12.2024
8	BROD - FC	01.12.2024
9	BROD - FC	01.12.2024
10	BROD - FC	01.12.2024
11	BROD - FC	01.12.2024
12	BROD - FC	01.12.2024
13	BROD - FC	01.12.2024
14	BROD - FC	01.12.2024
15	BROD - FC	01.12.2024
16	BROD - FC	01.12.2024
17	BROD - FC	01.12.2024
18	BROD - FC	01.12.2024
19	BROD - FC	01.12.2024
20	BROD - FC	01.12.2024
21	BROD - FC	01.12.2024
22	BROD - FC	01.12.2024
23	BROD - FC	01.12.2024
24	BROD - FC	01.12.2024
25	BROD - FC	01.12.2024
26	BROD - FC	01.12.2024
27	BROD - FC	01.12.2024
28	BROD - FC	01.12.2024
29	BROD - FC	01.12.2024
30	BROD - FC	01.12.2024
31	BROD - FC	01.12.2024
32	BROD - FC	01.12.2024

Exhibit 285

	A	B	C	D
1	**Platform**	**Page URL**	**Ad/Content URL**	**Language**
2	TikTok	[REDACTED]	[REDACTED]	Romanian
3	TikTok	[REDACTED]	[REDACTED]	Romanian
4	TikTok	[REDACTED]	[REDACTED]	Romanian
5	TikTok	[REDACTED]	[REDACTED]	Romanian
6	TikTok	[REDACTED]	[REDACTED]	Romanian
7	TikTok	[REDACTED]	[REDACTED]	Romanian
8	TikTok	[REDACTED]	[REDACTED]	Romanian
9	TikTok	[REDACTED]	[REDACTED]	Romanian
10	TikTok	[REDACTED]	[REDACTED]	Romanian
11	TikTok	[REDACTED]	[REDACTED]	Romanian
12	TikTok	[REDACTED]	[REDACTED]	Romanian
13	TikTok	[REDACTED]	[REDACTED]	Romanian
14	TikTok	[REDACTED]	[REDACTED]	Romanian
15	TikTok	[REDACTED]	[REDACTED]	Romanian
16	TikTok	[REDACTED]	[REDACTED]	Romanian
17	TikTok	[REDACTED]	[REDACTED]	Romanian
18	TikTok	[REDACTED]	[REDACTED]	Romanian
19	TikTok	[REDACTED]	[REDACTED]	Romanian
20	TikTok	[REDACTED]	[REDACTED]	Romanian
21	TikTok	[REDACTED]	[REDACTED]	Romanian
22	TikTok	[REDACTED]	[REDACTED]	Romanian
23	TikTok	[REDACTED]	[REDACTED]	Romanian
24	TikTok	[REDACTED]	[REDACTED]	Romanian
25	TikTok	[REDACTED]	[REDACTED]	Romanian
26	TikTok	[REDACTED]	[REDACTED]	Romanian
27	TikTok	[REDACTED]	[REDACTED]	Romanian
28	TikTok	[REDACTED]	[REDACTED]	Romanian
29	TikTok	[REDACTED]	[REDACTED]	Romanian
30	TikTok	[REDACTED]	[REDACTED]	Romanian
31	TikTok	[REDACTED]	[REDACTED]	Romanian
32	TikTok	[REDACTED]	[REDACTED]	Romanian
33	TikTok	[REDACTED]	[REDACTED]	Romanian
34	TikTok	[REDACTED]	[REDACTED]	Romanian
35	TikTok	[REDACTED]	[REDACTED]	Romanian
36	TikTok	[REDACTED]	[REDACTED]	Romanian
37	TikTok	[REDACTED]	[REDACTED]	Romanian
38	TikTok	[REDACTED]	[REDACTED]	Romanian
39	TikTok	[REDACTED]	[REDACTED]	Romanian
40	TikTok	[REDACTED]	[REDACTED]	Romanian
41	TikTok	[REDACTED]	[REDACTED]	Romanian
42	TikTok	[REDACTED]	[REDACTED]	Romanian
43	TikTok	[REDACTED]	[REDACTED]	Romanian
44	TikTok	[REDACTED]	[REDACTED]	Romanian
45	TikTok	[REDACTED]	[REDACTED]	Romanian
46	TikTok	[REDACTED]	[REDACTED]	Romanian
47	TikTok	[REDACTED]	[REDACTED]	Romanian
48	TikTok	[REDACTED]	[REDACTED]	Romanian
49	TikTok	[REDACTED]	[REDACTED]	Romanian
50	TikTok	[REDACTED]	[REDACTED]	Romanian
51	TikTok	[REDACTED]	[REDACTED]	Romanian
52	TikTok	[REDACTED]	[REDACTED]	Romanian
53	TikTok	[REDACTED]	[REDACTED]	Romanian
54	TikTok	[REDACTED]	[REDACTED]	Romanian
55	TikTok	[REDACTED]	[REDACTED]	Romanian
56	TikTok	[REDACTED]	[REDACTED]	Romanian
57	TikTok	[REDACTED]	[REDACTED]	Romanian
58	TikTok	[REDACTED]	[REDACTED]	Romanian
59	TikTok	[REDACTED]	[REDACTED]	Romanian
60	TikTok	[REDACTED]	[REDACTED]	Romanian

	E	F
1	**Summary**	**Organisation flagging**
2	Manipulative content related to elections	BROD - FC
3	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
4	Manipulative content related to elections	BROD - FC
5	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
6	Manipulative content related to elections	BROD - FC
7	Manipulative content related to elections	BROD - FC
8	Misinformation about elections	BROD - FC
9	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
10	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
11	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
12	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
13	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
14	Manipulative content related to elections	BROD - FC
15	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
16	Manipulative content related to elections	BROD - FC
17	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
18	Misinformation about elections	BROD - FC
19	Manipulative content related to elections	BROD - FC
20	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
21	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
22	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
23	Manipulative content related to elections	BROD - FC
24	Manipulative content related to elections	BROD - FC
25	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
26	Manipulative content related to elections	BROD - FC
27	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
28	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
29	Manipulative content related to elections	BROD - FC
30	Manipulative content related to elections	BROD - FC
31	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
32	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
33	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
34	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
35	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
36	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
37	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
38	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
39	Manipulative content related to elections	BROD - FC
40	Manipulative content related to elections	BROD - FC
41	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
42	Manipulative content related to elections	BROD - FC
43	Manipulative content related to elections	BROD - FC
44	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
45	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
46	Manipulative content related to elections	BROD - FC
47	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
48	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
49	Manipulative content related to elections	BROD - FC
50	Misinformation about elections	BROD - FC
51	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
52	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
53	Misinformation about elections	BROD - FC
54	Manipulative content related to elections	BROD - FC
55	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
56	Manipulative content related to elections	BROD - FC
57	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
58	Misinformation about elections	BROD - FC
59	Misinformation about elections	BROD - FC
60	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC

	G
1	**Reporting date**
2	03.12.2024
3	03.12.2024
4	03.12.2024
5	03.12.2024
6	03.12.2024
7	03.12.2024
8	03.12.2024
9	03.12.2024
10	03.12.2024
11	03.12.2024
12	03.12.2024
13	03.12.2024
14	03.12.2024
15	03.12.2024
16	03.12.2024
17	03.12.2024
18	03.12.2024
19	03.12.2024
20	03.12.2024
21	03.12.2024
22	03.12.2024
23	03.12.2024
24	03.12.2024
25	03.12.2024
26	03.12.2024
27	03.12.2024
28	03.12.2024
29	03.12.2024
30	03.12.2024
31	03.12.2024
32	03.12.2024
33	03.12.2024
34	03.12.2024
35	03.12.2024
36	03.12.2024
37	03.12.2024
38	03.12.2024
39	03.12.2024
40	03.12.2024
41	03.12.2024
42	03.12.2024
43	03.12.2024
44	03.12.2024
45	03.12.2024
46	03.12.2024
47	03.12.2024
48	03.12.2024
49	03.12.2024
50	03.12.2024
51	03.12.2024
52	03.12.2024
53	03.12.2024
54	03.12.2024
55	03.12.2024
56	03.12.2024
57	03.12.2024
58	03.12.2024
59	03.12.2024
60	03.12.2024

	A	B	C	D
61	TikTok	[REDACTED]	[REDACTED]	Romanian
62	TikTok	[REDACTED]	[REDACTED]	Romanian
63	TikTok	[REDACTED]	[REDACTED]	Romanian
64	TikTok	[REDACTED]	[REDACTED]	Romanian
65	TikTok	[REDACTED]	[REDACTED]	Romanian
66	TikTok	[REDACTED]	[REDACTED]	Romanian
67	TikTok	[REDACTED]	[REDACTED]	Romanian
68	TikTok	[REDACTED]	[REDACTED]	Romanian
69	TikTok	[REDACTED]	[REDACTED]	Romanian
70	TikTok	[REDACTED]	[REDACTED]	Romanian
71	TikTok	[REDACTED]	[REDACTED]	Romanian
72	TikTok	[REDACTED]	[REDACTED]	Romanian
73	TikTok	[REDACTED]	[REDACTED]	Romanian
74	TikTok	[REDACTED]	[REDACTED]	Romanian
75	TikTok	[REDACTED]	[REDACTED]	Romanian
76	TikTok	[REDACTED]	[REDACTED]	Romanian
77	TikTok	[REDACTED]	[REDACTED]	Romanian
78	TikTok	[REDACTED]	[REDACTED]	Romanian
79	TikTok	[REDACTED]	[REDACTED]	Romanian
80	TikTok	[REDACTED]	[REDACTED]	Romanian
81	TikTok	[REDACTED]	[REDACTED]	Romanian
82	TikTok	[REDACTED]	[REDACTED]	Romanian
83	TikTok	[REDACTED]	[REDACTED]	Romanian
84	TikTok	[REDACTED]	[REDACTED]	Romanian
85	TikTok	[REDACTED]	[REDACTED]	Romanian
86	TikTok	[REDACTED]	[REDACTED]	Romanian
87	TikTok	[REDACTED]	[REDACTED]	Romanian
88	TikTok	[REDACTED]	[REDACTED]	Romanian
89	TikTok	[REDACTED]	[REDACTED]	Romanian
90	TikTok	[REDACTED]	[REDACTED]	Romanian
91	TikTok	[REDACTED]	[REDACTED]	Romanian
92	TikTok	[REDACTED]	[REDACTED]	Romanian
93	TikTok	[REDACTED]	[REDACTED]	Romanian
94	TikTok	[REDACTED]	[REDACTED]	Romanian
95	TikTok	[REDACTED]	[REDACTED]	Romanian

	E	F
61	Manipulative content related to elections	BROD - FC
62	Manipulative content related to elections	BROD - FC
63	Manipulative content related to elections	BROD - FC
64	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
65	Misinformation about elections	BROD - FC
66	Manipulative content related to elections	BROD - FC
67	Manipulative content related to elections	BROD - FC
68	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
69	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
70	Manipulative content related to elections	BROD - FC
71	Manipulative content related to elections	BROD - FC
72	Manipulative content related to elections	BROD - FC
73	Misinformation about elections	BROD - FC
74	Manipulative content related to elections	BROD - FC
75	Manipulative content related to elections	BROD - FC
76	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
77	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
78	Manipulative content related to elections	BROD - FC
79	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
80	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
81	Manipulative content related to elections	BROD - FC
82	Misinformation about elections	BROD - FC
83	Manipulative content related to elections	BROD - FC
84	Manipulative content related to elections	BROD - FC
85	Manipulative content related to elections	BROD - FC
86	Misinformation about elections	BROD - FC
87	Misinformation about elections	BROD - FC
88	Manipulative content related to elections	BROD - FC
89	Account flagged for unmarked electoral promotion (content not labeled as electoral promotion)	BROD - FC
90	Misinformation about elections	BROD - FC
91	Misinformation about elections	BROD - FC
92	Manipulative content related to elections	BROD - FC
93	Misinformation about elections	BROD - FC
94	Fake information about the identity of Elena Lasconi and Eugen Tomac	BROD - FC
95	Propaganda for Calin Georgescu	BROD - FC

	G
61	03.12.2024
62	03.12.2024
63	03.12.2024
64	03.12.2024
65	03.12.2024
66	03.12.2024
67	03.12.2024
68	03.12.2024
69	03.12.2024
70	03.12.2024
71	03.12.2024
72	03.12.2024
73	03.12.2024
74	03.12.2024
75	03.12.2024
76	03.12.2024
77	03.12.2024
78	03.12.2024
79	03.12.2024
80	03.12.2024
81	03.12.2024
82	03.12.2024
83	03.12.2024
84	03.12.2024
85	03.12.2024
86	03.12.2024
87	03.12.2024
88	03.12.2024
89	03.12.2024
90	03.12.2024
91	03.12.2024
92	03.12.2024
93	03.12.2024
94	03.12.2024
95	03.12.2024

Section VIII:
Other Exhibits

Exhibit 286

MALIGN USE OF ALGORITHMIC AMPLIFICATION OF TERRORIST AND VIOLENT EXTREMIST CONTENT: RISKS AND COUNTERMEASURES IN PLACE

Authored by Redacted

Radicalisation Awareness Network
RAN Policy Support

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_005601

LEGAL NOTICE

This document has been prepared for the European Commission however it reflects the views only of the authors, and the European Commission is not liable for any consequence stemming from the reuse of this publication. More information on the European Union is available on the Internet (http://www.europa.eu).

Luxembourg: Publications Office of the European Union, 2021

© European Union, 2021

The reuse policy of European Commission documents is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/).This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

For any use or reproduction of elements that are not owned by the European Union, permission may need to be sought directly from the respective rightholders.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_005602

EXECUTIVE SUMMARY

This paper investigates the malign use of algorithmic amplification of terrorist and violent extremist content, as well as of so-called borderline or fringe content (i.e. content that is not illegal but may be harmful and exploited by extremist actors). The focus of this research is on understanding the risks linked to this phenomenon and the countermeasures in place, in order to identify gaps and challenges in addressing it. The study investigated the following research question: **how is algorithmic amplification used to promote terrorist and violent extremist content and how can the phenomenon be addressed?**

To answer this, five research sub-questions were investigated by means of a review of academic and grey literature, and expert and stakeholder interviews.

1. What is algorithmic amplification and how is it or can it be used to promote terrorist and violent extremist content?
2. What is the extent of the current threat posed by the potential misuse of such algorithms?
3. What is the role of the private sector in addressing this challenge?
4. What actions have already been taken by private-sector actors to address the phenomenon of algorithmic amplification to promote terrorist and violent extremist content?
5. What gaps, opportunities and challenges should policymakers address and consider?

In the context of this paper, algorithmic amplification is understood as the process of automated online content promotion with the intention of driving engagement. This may include content promoting terrorist and violent extremist views, disinformation and conspiracy theories aimed at legitimising extremist or terrorist content. Furthermore, it may be susceptible to human exploitation, where trending content such as hashtags or imagery is used to heighten the visibility of potentially malign content.

While terrorist and extremist groups have been able to exploit this technology in the past, current content moderation policies have made it more challenging for malicious actors, as overt terrorist and violent extremist content tends to be identified and removed effectively, at least on major online platforms. This has resulted in a shift in the strategies used by malicious actors. On the one hand, terrorist and violent extremist groups have adopted innovative techniques to host and share their content through alternative platforms, which are not as regulated as mainstream ones but have a more limited reach. On the other hand, terrorist and extremist groups have exploited the amplification of fringe content such as hate speech or conspiracy theories (which is potentially harmful but not illegal), reaching wider audiences. It is difficult to assess the extent of the threat posed by the exploitation of recommendation algorithms for fringe content, as there is only anecdotal evidence on the impact that this kind of content has in encouraging radicalisation and/or violent behaviour.

Inadvertent amplification of terrorist and violent extremism content can happen as a result of platforms' recommender algorithms promoting content with high levels of engagement. In some instances, malicious actors, including terrorist and violent extremist groups, are able to take advantage of such dynamics. More broadly, terrorist and extremist groups appear to have developed approaches and techniques to manipulate online platforms through deceitful strategies to support their influence operations. However, it is challenging to isolate these strategies from

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005603

the online ecosystem as a whole and to determine where inadvertent and deceitful amplification can be separated from one another.

In this context, private actors such as tech companies, social media platforms and online marketplaces have come under increased scrutiny to tackle this phenomenon and ensure that recommendation algorithms used on their platforms do not contribute to the dissemination of violent extremist and terrorist content. The debate on the role that private-sector actors should play focuses on whether and how the amplification of fringe content can be limited. As this debate is ongoing, existing measures identified by the study team reflect the variety of potential approaches to content moderation and to tackling the malign use of algorithmic amplification: (1) legal or regulatory measures; (2) technical measures to limit dissemination of extremist and/or harmful content; (3) technical measures to counter the potential impact of extremist and/or harmful content; (4) collaborative initiatives bringing together public and private actors; and (5) civil society initiatives.

The initiatives and discussion identified around these approaches have highlighted key challenges with regard to not only algorithmic amplification, but also the wider issue of online content moderation. These include: (1) challenges linked to a lack of understanding of how online platforms work; (2) a fragmented legal framework on issues of content moderation and censorship in the EU and globally; and (3) challenges related to issues of freedom of speech.

Based on these challenges, a number of recommended ways forward are proposed for the Commission and Member States:

- adopt a systemic approach to tackling online dissemination of extremist content;
- advance understanding of algorithmic amplification and its impacts through knowledge-oriented initiatives
- support the implementation of technological solutions through the definition of a framework for content moderation;
- investigate alternative legal measures to those targeting amplification through independent research;
- further invest in digital skills and capabilities for law enforcement to foster collaboration with platforms, including at national level.

KEYWORDS

Algorithmic amplification, algorithms, online promotion of extremist content, radicalisation, violent extremism, terrorism.

ABBREVIATIONS

ADL	Anti-Defamation League

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005604

CT	Counter-terrorism
DE	German
EFTA	European Free Trade Association
ES	Spain
EU	European Union
EUIF	European Union Internet Forum
FR	France
GIFCT	Global Internet Forum to Counter Terrorism
IRU	Internet Referral Unit
ISIS	Islamic State of Iraq and Syria
LSSI	Law on Information Society Services and Electronic Commerce
MS	Member States
NAACP	National Association for the Advancement of Colored People
OCCI	Online Civil Courage Initiative
OCLCTIC	Central Office for the Fight against Crime Relating to Information and Communication Technology
RAN	Radicalisation Awareness Network
T-CAP	Terrorist Content Analytics Platform
ToS	Terms of Service

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005605

INTRODUCTION

STUDY CONTEXT AND SCOPE

In recent years, terrorist and extremist groups have made increasing use of social media to disseminate their propaganda. As outlined in a recent EU Counter-Terrorism (CT) Coordinator report, these malicious actors have shown a detailed understanding of techniques and strategies to leverage online tools in order to reach as wide an audience as possible[1]. Algorithms used by online platforms have therefore come under increased scrutiny and criticism, as their role in disseminating terrorist and violent extremist content, as well as disinformation, hate speech and conspiracy theories is being discussed[2]. In this context, the technology of algorithmic amplification has entered the current policy discourse on how to tackle online radicalisation.

This paper investigates the malign use of algorithmic amplification for terrorist and violent extremist content, as well as so-called borderline or fringe content (which is not illegal but may be harmful and/or exploited by extremist actors). In doing so, this paper considers how regulatory measures to block terrorist and violent extremist content online have led to a strategic shift from malicious actors towards focusing on the dissemination of borderline or fringe content. It also provides an overview of the ongoing policy debate on the topic, and explores the extent of existing evidence on the potential impact that the algorithmic amplification of extremist and fringe content has on users' journeys towards radicalisation. The overall focus of this paper is on understanding the risks linked to this phenomenon, and the countermeasures that are in place, in order to identify gaps and challenges in addressing it.

STUDY APPROACH AND METHODOLOGY

The study sought to answer the following research question: how is algorithmic amplification used to promote terrorist and violent extremist content and how can the phenomenon be addressed? To answer this question, five research sub-questions were investigated.

1. What is algorithmic amplification and how is it or can it be used to promote terrorist and violent extremist content?
2. What is the extent of the current threat posed by the potential misuse of such algorithms?
3. What is the role of the private sector in addressing this challenge?
4. What actions have already been taken by private-sector actors to address the phenomenon of algorithmic amplification to promote terrorist and violent extremist content?
5. What gaps, opportunities and challenges should policymakers address and consider?

---

[1] EU Counter-Terrorism Coordinator (2020). 'The role of algorithmic amplification in promoting violent and extremist content and its dissemination on platforms and social media'.

[2] Stark and Stegmann (2020), 'Are Algorithms a threat to democracy?' Institute for Strategic Dialogue (2021). Recommended reading: Gosh and Scott (2018), #Digitaldeceit; Center for Countering Digital Hate (2020), Malgorithm; Krasofomski-Jones (2020), Everything in moderation.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005606

The study entailed a review of academic and grey literature on the focus subject. Specifically, literature was reviewed in English for selected sources on (1) algorithmic amplification and its malign use to promote terrorist and violent extremism content online; (2) the role of the private sector and existing initiatives to address this phenomenon; (3) gaps, opportunities and challenges that policymakers should consider when addressing this phenomenon. Additional sources were reviewed as a result of snowballing from those already identified, and based on specific recommendations that emerged from interviews conducted by the study team. In parallel to literature review activities, the study team conducted five expert and stakeholder interviews with academics, law enforcement experts and data analysis specialists. The overview of sources consulted along with details on the interview approach and respondents can be found in Annex A.

CAVEATS AND LIMITATIONS

Some caveats and limitations should be noted when reviewing this paper.

- **Findings are based on open-source documentation and academic sources.** This paper relies on academic sources, as well as on open-source documentation that is available in English. As such, it may not provide a comprehensive overview of provisions, data and measures that are not available in the public domain due to their confidential or sensitive nature. With regard to policies and measures put in place by online platforms, for instance, this paper considers only data available in the public domain and/or discussed in academic literature. The same applies for measures established by public authorities, including initiatives by law enforcement agencies.

- **Findings are based on a limited number of stakeholder and expert consultations**. While interviews conducted by the study team provided extensive insights into the functioning of algorithmic amplification, and its associated risks and countermeasures, it should be noted that time and resource constraints did not allow the study team to conduct wider stakeholder consultations. The study team sought to mitigate this by engaging with different types of stakeholders, including perspectives from law enforcement, online platforms, academics and civil society (please see Annex A for further details). Unless otherwise noted in the paper, statements linked to a majority view or multiple interviewees consulted represent the perspective of stakeholders and experts from different sectors and organisations. It should nonetheless be noted that experts consulted were asked to contribute their individual perspectives and insights, rather than to represent those of their organisations or sector.

- **Scope of recommendations**. The paper endeavours to identify gaps and challenges, and propose possible solutions to challenges stemming from algorithmic amplification. However, further work is needed to build the evidence base necessary to understand the radicalisation and extremism threat associated with algorithmic amplification. In addition, further research is needed to assess the effectiveness of measures discussed in this paper. These evidence and knowledge gaps limit the extent to which the study team could formulate robust and evidence-based recommendations on these subjects.

REPORT STRUCTURE

In addition to this section, this paper contains four additional sections.

- Section 2 provides an overview of the malign use of algorithmic amplification of terrorist and violent extremist content.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005607

- Section 3 discusses the role of the private sector in preventing the malign use of algorithmic amplification, and provides an overview of countermeasures mapped by the study team.
- Section 4 provides an overview of gaps, opportunities and challenges in addressing the malign use of algorithmic amplification.
- Section 5 synthesises the key findings and proposes a set of policy recommendations.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005608

OVERVIEW OF THE MALIGN USE OF ALGORITHMIC AMPLIFICATION OF TERRORIST AND VIOLENT EXTREMIST CONTENT

CHALLENGES IN UNDERSTANDING ALGORITHMIC AMPLIFICATION

Recent discourse on radicalisation in online spaces refers to algorithmic amplification as a driver of the visibility of extreme content. However, it is important to note the challenges in understanding the technology behind the term for non-computer scientists, especially with regard to understanding its malign use and exploitability. An algorithm can be understood as any set of sequential instructions that automate how computers perform certain functions[3]. This includes a multitude of standardised problem-solving processes and is an integral aspect of all online interactions. The term 'algorithmic amplification' therefore describes any process that increases the circulation rate of any particular information or content online[4].

The literature identifies three main categories of algorithms, all of which contain so-called amplification functions. Firstly, **search algorithms** interface with search functions on platforms such as Google, social media platforms, and online marketplaces such as Amazon or Etsy. User-specific factors such as geographic location and language determine how results are ranked to avoid irrelevant content. Secondly, **recommender algorithms** have been the main focus of recent studies on extremist online content[5]. These algorithms make automated, personalised recommendations based on previous content engagements for user groups. With increased use, recommender algorithms refine results to match the user's persona and preferences. Lastly, **advertising technology algorithms** target users through known interests to expose them to relevant products, and are a key source of income for digital platforms. This algorithmic function has been heavily criticised for attempted influence on voting behaviours[6]. All three categories of algorithms rely on both passive and active behavioural indicators, including gender, age, location, time spent viewing content, and likes or shares, to individualise the user experience and drive further engagement[7].

Several interview respondents and literature consulted for this paper highlight the widespread use of algorithmic amplification as a structural feature of digital platforms that is essential to enable, along with other technologies leveraging machine learning and artificial intelligence, a tailored and more structured user experience[8]. Should the algorithmic recommendation feature be removed entirely, users would be presented with unfiltered and unstructured content that would complicate the practical navigation of search platforms, online marketplaces and social media. For example, search results on e-commerce platforms would offer irrelevant products when not adjusted for the user's age, language and geographic location. In this context, it was underlined that current applications of the term 'algorithmic amplification' do not accurately capture how widespread the technology is, nor the standardised and often useful ways in which it functions. Interviews conducted by the

[3] RAND Europe interview.

[4] Keller (2021), 'Amplification and its Discontents', Columbia University.

[5] Institute for Strategic Dialogue (2021a); Center for Countering Digital Hate (2020).

[6] Ghosh and Scott (2018).

[7] Hutchinson et al. (2021), 'Mobilizing extremism online: comparing Australian and Canadian right-wing extremist groups on Facebook'; RAND Europe interviews.

[8] RAND Europe interviews. Bellasio et al (2021). Ignatidou (2019).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005609

study team highlighted the need to ensure a common understanding of the technology and its uses when leveraging the term 'algorithmic amplification'. In this context, some stakeholders suggest the use of alternative terms such as 'content personalisation', which captures the tool's multifaceted uses without implying proactive use for specific purposes by digital platforms or other actors.[9]

For the purposes of this paper, we understand **algorithmic amplification to be the process of automated content promotion online for engagement-driving purposes**. This may include content promoting terrorist and violent extremist views, as well as disinformation and conspiracy theories. Furthermore, it may be susceptible to human exploitation, where trending content such as hashtags or imagery is used to heighten the visibility of potentially malign content.

OVERVIEW OF THE MALIGN USE OF ALGORITHMIC AMPLIFICATION

A - Algorithmic amplification of terrorist and violent extremist content

The algorithmic amplification of terrorist and violent extremist content can occur through widespread integration of this content into public discourse. Examples of this phenomenon have been seen in recent years, when there have been instances of designated terrorist content, specifically videos, that reached millions of people through amplification on mainstream social media platforms.

A number of notable instances of this phenomenon are discussed in the literature. One such example is that of the Christchurch terrorist attack video. The lone terrorist conducting the attack live-streamed his actions, which resulted in the fatal shooting of 51 people. Originally, the link was posted to a right-wing extremist group on the social media platform 8chan, from which it was subsequently mass-shared on Facebook.[10] In this case, algorithmic amplification of extremist content occurred as a result of the mainstream media discourse online and offline absorbing and amplifying violent content that was of broad public interest or relevance.

Additional examples include the 2015 Islamic State (ISIS) propaganda video showing the execution of Jordanian pilot Muath al-Kaseasbeh, who was burned alive. The video was shared widely online and US television channel Fox News showed it in its entirety as a condemnation of ISIS acts.[11] Similarly, in 2019, four ISIS-affiliated attackers murdered two Scandinavian women at a campsite in the Atlas Mountains and shared a video of the attack. This was widely distributed and sent to the victims' families by users expressing outrage at the acts – many of whom were unaware that the video was classified as terrorist content.[12] **These examples highlight how algorithmic content recommendations can be inadvertently exacerbated by non-malicious actors, when terrorist and violent extremist content enters the mainstream public discourse and its offline and online ecosystems.**

---

[9] RAND Europe interviews; Tech Against Terrorism (2021a), 'Position paper Content personalisation and the online dissemination of terrorist and violent extremist content'.

[10] Macklin (2019), 'The Christchurch Attacks: Livestream Terror in the Viral Video Age'.

[11] BBC (2015), 'Fox News explains why it showed Jordan pilot video'.

[12] BBC (2019), 'Morocco hikers: Three get death penalty for Scandinavian tourist murders.'

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005610

Indeed, before the recent implementation of content removal frameworks and legislation prohibiting violent and extremist online content, prominent terrorist groups such as ISIS used major social media platforms for recruitment and propaganda dissemination purposes. In 2014, ISIS created an Arabic language Android application called The Dawn of Glad Tiding to post news and propaganda content from their territorial strongholds in Syria and Iraq[13]. The app had a mass-sharing feature allowing users to generate and post messages on Twitter automatically. This enabled extensive amplification of content, with up to 40 000 tweets posted in one day[14]. This was also combined with the use of bots to spread large amounts of propaganda through social media platforms and influence their ranking on search results and through recommendations[15].

In addition to mass-sharing the auto-generated content produced by ISIS central communication affiliates, Veilleux-Lepage reports that the group also reportedly engaged in so-called 'hashtag hijacking'[16]. Hashtag hijacking involves using trending hashtags and adding them to unrelated content in order to increase its visibility, as algorithms will classify this content as relevant in the context of the hashtags used. This allows extremist content to quickly gain traction and reach new audiences that would not otherwise be exposed it. Examples of ISIS hashtag hijacking include the 2014 Scottish Independence Referendum and the World Cup. In both instances, members of ISIS' communication unit used trending hashtags to promote propaganda videos featuring kidnapping victims and foreign fighters in English and Arabic[17].

However, widespread amplification of designated terrorist content on platforms such as Facebook, Twitter and YouTube has decreased significantly in response to increased content moderation efforts[18]. Legislation regulating online speech inciting violence and hatred against protected groups, as well as technical and collaborative initiatives (see Table 3-1 in Section 3), have resulted in large platforms suspending users who spread overtly hateful or violent content. Content moderation efforts have intensified significantly, particularly among major platforms such as Facebook, Twitter and Google[19]. For example, these platforms have implemented mechanisms that combine automated machine learning and human assessment to detect ISIS-related content[20].

In this context, Facebook claims that 99.6% of the designated terrorist content (e.g. related to ISIS or its affiliates) removed in the first quarter of 2021 was found and flagged before it was reported by users[21]. In addition, YouTube tracks the total number of items removed on a bimonthly basis, nearing 1 million per month. Amazon disclosed its removal of 92 books on Holocaust denial in January 2021[22]. Section 3 of this paper provides further details on exiting initiatives to combat extremist content online, and while challenges remain – as discussed in

---

[13] Veilleux-Lepage, Y. (2016), 'Paradigmatic Shifts in Jihadism in Cyberspace: The Emerging Role of Unaffiliated Sympathizers in Islamic State's Social Media Strategy'.

[14] Vitale and Keagle (2014), 'A Time to Tweet, as well as a Time to Kill: ISIS' Project of Power in Iraq and Syria', Institute for National Strategic Studies.

[15] Lieberman, A. V. (2017), 'Terrorism, the Internet, and Propaganda: A Deadly Combination'; United Nations Office of Counter-Terrorism (UNOCT) (2021), 'Algorithms and Terrorism: The Malicious Use of Artificial Intelligence for Terrorist Purposes'.

[16] Veilleux-Lepage, Y. (2016).

[17] Veilleux-Lepage, Y. (2016).

[18] RAND Europe interview; Bossong, R. (2021), 'The Next Steps for EU Counterterrorism Policy. Evolving Threats of Jihadism, Right-wing Extremism, and Transatlantic Cooperation'; Tech Against Terrorism (2021a).

[19] RAND Europe interview.

[20] Gorwa et al. (2020).

[21] Facebook (2021), 'Dangerous Organizations: Terrorism and Organized Hate'.

[22] Stark and Stegmann (2020); YouTube (2021), 'Transparency Report'; Institute for Strategic Dialogue (2021a); RAND Europe interviews.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005611

Section 4 of the paper – several of these have resulted in promising successes when it comes to removing designated terrorist and illegal content.

As a result of the initiatives and efforts discussed above and in Section 3 of this paper, it has become increasingly difficult for designated terrorist groups to maintain a stable, visible presence on mainstream online platforms[23]. Instead, terrorist and violent extremist groups appear to have shifted tactics to disseminate their propaganda and reach wide audiences. One interviewee highlighted the use of smaller, encrypted, functional platforms such as content storage and mirroring platforms[24]. Importantly, a significant proportion of such platforms do not use algorithmic amplification technologies[25]. Content is disseminated through alternative mechanisms, including private messaging platforms, which are often encrypted, making it difficult to monitor and analyse dissemination patterns. In addition, groups are using increasingly sophisticated techniques, such as blurring terrorist logos in images, to avoid detection by content-trawling algorithms that moderators use to identify illegal content[26].

Other examples, notably used by ISIS supporters, include altering texts by placing commas and full stops in the middle of words to ensure that keywords cannot be recognised; hacking existing accounts not previously flagged for sharing illegal content; or using harmless video content as an introduction to violent propaganda videos[27]. Increasingly elaborate techniques used by terrorists and violent extremists online therefore mainly focus on avoiding content moderation rather than amplifying content[28].

B - Algorithmic amplification of borderline or fringe content

This recent shift in terrorist content removal has prompted a strategic shift in violent extremist groups' use of online spaces[29]. In addition to extremist groups using innovative tactics to avoid online detection, **there are increasing concerns over the algorithmic amplification of so-called fringe or grey-zone content**. This content, also informally referred to as **'lawful but awful' or 'legal but potentially harmful'**, is seen as a key challenge for policymakers and content moderators[30].

Fringe content is subject to the same algorithms as any other legal content, personalising recommendations to certain users and amplifying popular elements. Current research highlights that **the sensationalism of some fringe content, including elements of violence or**

---

[23] RAND Europe interviews.

[24] RAND Europe interviews; Tech Against Terrorism (2019), 'ISIS use of smaller platforms and the DWeb to share terrorist content'.

[25] Tech Against Terrorism (2021a); Whittaker, J. et al. (2021), 'Recommender systems and the amplification of extremist content'.

[26] Tech Against Terrorism (2021a); Bipartisan Policy Center, Task Force on Terrorism and Ideology (2018), 'Digital Counterterrorism: Fighting Jihadists Online'.

[27] United Nations Office of Counter-Terrorism (UNOCT) (2021), 'Algorithms and Terrorism: The Malicious Use of Artificial Intelligence for Terrorist Purposes'; Ayad, M. 'Islamic State terrorist propaganda is going viral on Facebook.'

[28] RAND Europe interviews.

[29] Burwell (2021), 'Free speech and online content: What can the US learn from Europe?'

[30] RAND interviews (this term was mentioned in several interviews conducted by the study team) and in the following report: Techdirt (2021), 'Regulating Amplification Is A Lot Harder Than You Think'. Meanwhile, the EU CT Coordinator 2020 report referred to the spread of 'legal but potentially harmful content'.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005612

hateful discourse, triggers emotional responses that prompt users to interact with the content by commenting on or sharing posts[31].

Several studies have identified how this type of content gets recommended through content personalisation on different platforms. For example, a longitudinal analysis on YouTube has shown that provoking videos featuring conspiracy theories lead to high engagement and are therefore further recommended to users. The most prominent conspiracy theories detected were broadly categorised into alternative scientific/historical facts, such as flat Earth theory, religious cults, and political conspiracies such as QAnon and the US 2020 election fraud campaigns. The latter were reported by American intelligence services as likely to motivate acts of violent extremism, particularly with regards to QAnon and Pizzagate, which are theories related to the New World Order and the claim that the UN is an organisation run by global elites to establish global rule[32].

While the study by Faddoul et al. (2020) shows that the recommendation algorithm no longer promotes more extreme videos as it had in the past, the authors highlight that clicking on a conspiratorial video will lead to the recommendation of similar content[33]. Meanwhile, a study of Amazon's recommendation algorithm concluded that the algorithm may actively promote books by extremist authors, auto-completes searches with conspiracy-related suggestions, and recommends books promoting disinformation and white nationalism to users who searched for content using key terms associated with conspiracy theories. These results were prioritised over factual, nuanced alternatives for those users[34]. Furthermore, recent research suggest similar results when considering books related to anti-Semitism, nativism, Nazism and white nationalism on Amazon – with a particular concern for self-published books as the pre-publication review process focuses solely on illegal or indecent content.[35] Similarly, a qualitative study on vaccine misinformation on Instagram showed that the platform's recommender algorithm promoted false and misleading content, rather than factual content from health authorities, after minimal indications of user interest. Furthermore, topics such as US 2020 election fraud, anti-Semitism, and the QAnon conspiracy theory were introduced to users following anti-vaccine channels[36].

Research also highlighted that extremists, like many content creators and distributors, take advantage of online infrastructure, including the functioning of recommendation algorithms, and base their content dissemination strategy on them. This includes choosing adequate words, titles and thumbnails to promote a video, for instance, both for the purpose of reaching the widest possible audience and avoiding content moderation37. Sources also suggest that certain extremist groups leverage the mechanisms for amplification to spread their ideology by hijacking proxy issues. This essentially means that they use current societal trends or divisive issues that are widely discussed in mainstream discourse as a platform to spread their own narratives to wider audiences[38]. In the context of the Covid-19 pandemic,

---

[31] Faddoul, M., Chaslot, G., Farid, H. (2020), 'A longitudinal analysis of YouTube's promotion of conspiracy videos'; Center for Countering Digital Hate (2020).

[32] Faddoul, M., Chaslot, G., Farid, H. (2020).

[33] Faddoul, M., Chaslot, G., Farid, H. (2020).

[34] Institute for Strategic Dialogue (2021a);

[35] Kofman, A., Tseng, F. and Weigel, M. (2020) 'The Hate Store: Amazon's Self-Publishing Arm Is a Haven for White Supremacists')

[36] Center for Countering Digital Hate (2020).

[37] RAND Europe interviews; Tech Against Terrorism (2021a); Bipartisan Policy Center, Task Force on Terrorism and Ideology (2018).

[38] RAND Europe interviews; Davey et al. (2019), 'The mainstreaming of far-right extremism online and how to counter it. A case study on UK, US and French elections.'

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005613

this has manifested as vaccine scepticism escalating into 5G conspiracies that now dominate existing borderline and disinformation communities online, such as QAnon groups[39].

Moreover, **both extremist and fringe groups are increasingly mobilising on small and medium-sized platforms. Labelled as 'alt-tech' and 'decentralised web' platforms in the literature, extremist groups consider such platforms as 'beacons of free speech'**[40]. These platforms, identified by interview respondents, include Parlor, Gab, Telegram, DLife, 4chan and Reddit[41]. Extremist and disinformation content is strategically disseminated among key actors across these platforms and then launched on mainstream platforms (such as Facebook and Twitter) by a combination of authentic followers and bots, to exploit amplification mechanisms[42]. Davey et al. (2019) have highlighted that specific content such as memes or conspiracy theories are spread on mainstream social media at critical moments (e.g. during a political campaign) by activists, fake accounts and bots in order to amplify their reach[43]. Examples discussed in existing literature include white supremacist channels on Telegram amplifying Covid-19 disinformation to attract an 800% increase in followers, and far-right groups harnessing the Reopen Movement's anti-lockdown sentiment to sustain their messaging[44].

These examples highlight how the **algorithmic amplification of so-called fringe content is, to some extent, used to spread hateful narratives, conspiracy theories and disinformation, including by extremist groups.** Tackling the amplification of such content, by its very nature, raises important ethical questions about content moderation and limitations on free speech, as discussed in Section 4 of this report.

C - Inadvertent and deceitful algorithmic amplification of terrorist and violent extremist content

As discussed under sections A and B of this chapter, algorithmic amplification can facilitate the dissemination of designated terrorist and violent extremism content, as well as of borderline or fringe content, including some purposedly designed and disseminated by terrorist and violent extremist actors. This section discusses how recommender algorithms can amplify harmful content both inadvertently and as a result of deceitful strategies used by terrorists and extremists to manipulate this technology.

Inadvertent amplification of terrorist and violent extremism content can happen as a result of recommender algorithms promoting content with high levels of engagement. With increasingly strict content moderation rules, inadvertent amplification of terrorist content tends to happen as a result of such content entering the mainstream public discourse and being posted and shared by large volume of users, typically not intending to do so in support of the groups or content depicted.

---

[39] Thomas and Zhang (2020), 'ID2020, Bill Gates and the Mark of the Beast: how Covid-19 catalyses existing online conspiracy movements.'

[40] RAND Europe interviews.

[41] Interview respondent 04, interview by authors (6 July 2021); Erin Saltman, interview by authors (2 July 2021).

[42] Davey et al. (2019); Cox et al. (2021), 'COVID-19, Disinformation and Hateful Extremism'. Literature review report prepared for the Commission for Countering Extremism (CCE); Müller & Denner (2019), p. 9; Wardle & Derakshan (2018), p. 49 quoted in Stark and Stegmann, D. (2020).

[43] Davey et al. (2019).

[44] Crawford, (2020), 'Coronavirus and conspiracies: how the far right is exploiting the pandemic'.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005614

Examples of this phenomenon include the dissemination of the Christchurch terrorist attack video, the video showing the execution of Jordanian pilot Muath al-Kaseasbeh or the video of the Atlas Mountains attack.[45] When algorithmic amplification occurs for content that can be labelled as terrorist and/or violent extremist in nature (i.e. in violation of the Terms of Service of platforms), provisions and safeguards implemented by online platforms designed to ensure such content is taken down (e.g. by relying on content moderation mechanisms) intervene to halt its distribution. However, when this type of inadvertent amplification happens, due to the high levels of engagement and traction that such content can generate, amplification facilitates a wider circulation of materials, reaching a broader range of actors and audiences than otherwise possible for terrorist and violent extremist content. Nonetheless, **it is difficult to evaluate the extent to which malicious actors are able to take advantage of this phenomenon on a systematic basis** as these events appear to stem from specific incidents and to be circumscribed to individual videos or contents distributed by users not with propagandistic objectives. Similarly, it is difficult to ascertain the effectiveness of safeguard measures and response mechanisms in place due to limited information available on this type of incidents and on the functioning of recommender algorithms, particularly in response to the dissemination by a large volume of users of viral content of a violent or terrorist nature.

Meanwhile, content that is not openly violent or illegal is particularly likely to be subject to algorithmic amplification. This appears to be due to the fact that content with elements of violence, hate speech and/or conspiracy theories is likely to trigger emotional responses. Through this sensationalism, fringe or borderline content prompts high level of user engagement and therefore may get promoted through recommender algorithms. [46] Recent research has highlighted how content personalisation technologies on Youtube, Amazon, and Instagram inadvertently participate in amplifying the reach of content featuring conspiracy theories, disinformation, and far-right ideologies such as white nationalism or anti-Semitism. Such recommendations cover content that is either below the threshold or threading on the line of what can be identified as terrorist or violent extremist content, taking advantage of lack of clarity and uniformity in what can be labelled as extremist content for some ideologies and inclinations (e.g. for right-wing extremism).[47]

Terrorists and extremists have learnt to manipulate existing online infrastructure through deceitful strategies to maximise the reach of their material and influence efforts. This phenomenon adds further complexity to the issue of inadvertent amplification of extremist and terrorist content on mainstream online platforms. In particular, this is done by deploying technologies such a **bots, mass-sharing features** or **fake accounts** as well as tactics such as **hashtag hijacking**. Prior to the introduction of stricter content moderation rules, research has shown that terrorist groups, such as ISIS, introduced automated mass-sharing of Twitter posts, bots and hashtag hijacking (e.g. of the World Cup or Scottish Independence Referendum) in order to amplify their propaganda and reach new audiences through the manipulation of content recommendations.[48] Research also shows that similar tactics are now used to promote content which is not openly illegal but includes hateful discourses, violence and/or extremist theories. In particular, hashtag hijacking has been used to disseminate propaganda on right-wing extremism and/or conspiracy theories through various social media platforms.[49] Most recent research found that extremists on TikTok also include general hashtags related to current trends (e.g. Black Lives Matter; Republican;

---

[45] See Section A above and Macklin (2019), BBC (2015 & 2019)

[46] Faddoul, M., Chaslot, G., Farid, H. (2020); Wood (2017), 'Antisocial media and algorithmic deviancy amplification: Analysing the ID of Facebook's technological unconscious'.

[47] See Section B above and Faddoul, M., Chaslot, G., Farid, H. (2020); Institute for Strategic Dialogue (2021a); Kofman, A., Tseng, F. and Weigel, M. (2020); Center for Countering Digital Hate (2020).

[48] See Section A above and Vitale and Keagle (2014); Lieberman, A. V. (2017); Veilleux-Lepage, Y. (2016).

[49] See Section B above and RAND Europe interviews; Davey et al. (2019); Thomas and Zhang (2020)

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005615

Trump) in their posts to manipulate the recommender algorithm and ensure their content is disseminated to wide audiences.[50] Right-wing extremists have also been shown to exploit amplification mechanisms through the use of bots and fake accounts to deploy contents at critical moments, therefore maximising the reach of fringe and borderline content.[51]

Terrorists and extremists appear to combine approaches and methods to manipulate algorithmic amplification and recommender algorithms into broader deceitful tactics designed to support their influence operations via mainstream platforms. It should be noted that approaches targeting algorithmic amplification and recommender algorithms are usually part of wider strategies for content dissemination online, whereby terrorists and extremists make use of online infrastructure to their advantage. The manipulation of algorithmic amplification is one tenant of such strategies, which may also include the manipulation of search processes (e.g. by tagging extremist literature to obscure and niche categories on Amazon or by using misleading content to illustrate an account or the start of a video), as well as an increasing reliance on the use of smaller, alternative platforms.[52]

As regards exploiting algorithmic amplification, an integral part of terrorist and violent extremists strategies focuses on avoiding content moderation mechanisms and safeguards. This can include the blurring of terrorist logos to avoid automatic detection whilst leaving sufficient information for initiated eyes, as well as the deliberate introduction of altered text to avoid automated detection mechanisms.[53] Recent research focusing on the use of TikTok corroborates these findings by highlighting how the use of alternative spellings or similar words is leveraged by terrorists to bypass banned hashtags on the platform.[54]

Overall, it is also worth noting that whilst terrorist groups still employ mainstream platforms to disseminate content, there appears to have been a shift in the type of content spread through these by employing algorithmic amplification. In particular, research reviewed points to content disseminated via mainstream platforms in this manner to have less emphasis on promoting direct recruitment and propaganda, and instead to present more subtle content designed for influence purposes. In particular, on mainstream platforms, efforts by terrorist and extremist groups focus on fringe content which is not aimed at direct recruitment, but rather seeks to influence the public discourse, shift debates, and/or prepare the breeding ground for recruitment at a later stage. Indeed, as discussed in previous sections, direct recruitment and overt propaganda efforts by terrorist groups tend to be delivered through different means, especially on smaller platforms where the focus is not on amplification of content to a wide audience, but rather on organising and recruiting with smaller, receptive audiences while maintaining secrecy.[55]

The use of such tactics demonstrates the extent to which terrorists and extremists have developed competences to leverage and manipulate for their purposes online infrastructure, including mainstream platforms. In this context, the distinction between inadvertent and deceitful algorithmic amplification of terrorist and violent extremist content is one which is difficult to transpose from a conceptual perspective into the realities of the (online) ground. It appears that terrorists and extremists have learnt to develop and disseminate content

---

[50] Institute for Strategic Dialogue (2021c).

[51] See Section B above and Davey et al. (2019).

[52] Davey et al. (2019); Tech Against Terrorism (2021a); Bipartisan Policy Center, Task Force on Terrorism and Ideology (2018); Kofman, A., Tseng, F. and Weigel, M. (2020); RAND Europe interviews; Tech Against Terrorism (2021a); Bipartisan Policy Center, Task Force on Terrorism and Ideology (2018).

[53] United Nations Office of Counter-Terrorism (UNOCT) (2021;

[54] Institute for Strategic Dialogue (2021c).

[55] See sections A and B as well as RAND Europe interviews; Tech Against Terrorism (2019), Tech Against Terrorism (2021a); Whittaker, J. et al. (2021); Bipartisan Policy Center, Task Force on Terrorism and Ideology (2018); [55] United Nations Office of Counter-Terrorism (UNOCT) (2021); Ayad, M.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005616

triggering engagement and therefore further amplification by focusing on material which is difficult to directly identify as illegal, but which triggers emotional reactions and engagement.

This hypothesis, however, would need to be further investigated with direct access to relevant data. Yet, as discussed in greater detail under Chapter 4 of this paper, lack of transparency, access to, and understanding of proprietary recommender algorithms used by platforms further limits the ability to assess the impact of these tactics and of wider information operations conducted on mainstream platforms by extremist groups. **In particular, without access to data from platforms, it is challenging to isolate these strategies from the online ecosystem as a whole and to determine where inadvertent and deceitful amplification can be separated from one another.**

UNDERSTANDING AND PERCEPTION OF THE THREAT POSED BY MALIGN USE OF ALGORITHMIC AMPLIFICATION

Examples discussed under Section 2.2 highlight the extent to which algorithmic amplification can contribute, inadvertently or purposefully, to the spreading of terrorist and violent extremist content, as well as hate speech, conspiracy theories and disinformation. While the growing availability and dissemination of this type of content raises concerns[56], **understanding the impact of such content is challenging. This makes it difficult to assess the level of threat that this phenomenon represents.**

Several sources highlight the risk posed by selective exposure, or so-called filter bubbles and echo chambers, resulting from personalisation of online content[57]. For instance, in a review of Facebook pages for illegal fighting groups, Wood (2017) suggests that, through selective exposure, such pages may contribute to normalising violent behaviours and reinforcing violent codes of hypermasculinity and racist beliefs expressed through the illegal fight culture[58]. Faddoul et al. (2020) also point out the risk of 'reinforcing feedback loops' created by recommendations of videos featuring conspiracy theories on YouTube for users who consume them[59]. Miller (2018) highlights similar anecdotal evidence investigating online extreme right-wing radicalisation pathways, as recounted by extremists on an online forum.

These testimonies indicate the consumption of increasingly extreme content as a result of their own searches or recommendation algorithms[60]. In addition, others highlight risks that algorithmic amplification can pose to democracy when manipulated by malicious actors, whether foreign states or extremist groups, to encourage the mainstreaming of fringe narratives[61]. Considering different elections in the UK, the US and France, Davey et al. (2019) found that extremist activists increasingly coordinated campaigns to manipulate online conversations and spread alt-right narratives. While there is no robust evidence concerning the impact of such online activities, Davey et al. (2019) point to anecdotal evidence of the Pizzagate online conspiracy. This occurred during the 2016 US election campaign, when users alleged that a restaurant in Washington D.C. served as a child trafficking hub. The conspiracy quickly spread on social media and eventually led to a shooting at the restaurant by one individual who had consumed a large amount of related content online[62].

---

[56] EU Counter-Terrorism Coordinator (2020); Cox et al. (2021).

[57] Stark and Stegmann, D. (2020); Collins (2020), 'Jacinda Ardern calls for 'ethical algorithms' to combat online extremism. What this means.' The Conversation; Institute for Strategic Dialogue (2021a).

[58] Wood (2017).

[59] Faddoul et al. (2020).

[60] Miller (2018), 'McInnes, Molyneux, and 4chan: Investigating pathways to the alt-right'.

[61] Cox et al. (2021); Davey et al. (2019).

[62] Davey et al. (2019).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005617

However, literature and experts consulted by the study team emphasised that **it is challenging to link radicalisation pathways to the consumption of online content and to algorithmic amplification specifically**[63]. While anecdotal evidence might point to the role of content consumed and discovered online for specific individuals, **there is limited empirical evidence that such content can, in effect, lead to and correlate with violent behaviours and terrorist acts.** This is because factors, drivers and dynamics leading to radicalisation are complex and challenging to isolate from one another. These may include a combination of individual socio-economic, psychological, political, ideological and cultural dynamics both online and offline[64]. Interviews conducted by the study team highlighted the same challenges in identifying radicalisation factors, and specifically in isolating evidence of online radicalisation[65].

[63] RAND Europe interviews; Tech Against Terrorism (2021a); Ganesh and Bright (2019), 'Extreme Digital Speech: Contexts, Responses, and Solutions'.

[64] Ranstorp, M. (2016), 'The Root Causes of Violent Extremism. RAN Issue Paper'; Bellasio, J. et al. (2018), 'Counterterrorism evaluation: Taking stock and looking ahead'; RAND Europe interviews.

[65] RAND Europe interviews.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005618

THE ROLE OF THE PRIVATE SECTOR IN PREVENTING THE MALIGN USE OF ALGORITHMIC AMPLIFICATION

In light of examples of malicious and inadvertent amplification of terrorism and violent extremism content highlighted in Section 2, a debate has been sparked in recent years on the role of private-sector actors in preventing the malign use of algorithmic amplification.

Considering measures already in place for the removal of overtly terrorist content, **this debate now focuses mainly on the regulation of fringe content and on whether and how its amplification should be limited**. It has been argued that, while there might not be grounds for removing fringe content, platforms should be held accountable for the amplification of such content and, in particular, the removal of recommendation algorithms[66]. Although, as outlined under Section 2.1, the complete removal of algorithmic amplification would risk being detrimental to all users' experience of online platforms, several have argued for limiting amplification of fringe content only[67].

This links into the core question in this debate: attribution of responsibility for defining what constitutes fringe content. On the one hand, proponents of legal liability for amplification of harmful content argue that the commercial interests of private actors have prevailed, encouraging algorithmic amplification of content that pushes users to spend more time online. In this context, it is argued that stringent regulations are required to ensure that platforms are incentivised to limit this phenomenon[68]. These calls also illustrate what can be described as the privatisation of public spaces, whereby public discourse is increasingly occurring on private platforms. This is also reflected in Krasodomski-Jones's (2020) discussion of the risk of commercial interests subjugating public values[69].

On the other hand, criticism of the introduction of legal liability for amplifying fringe content highlights that this could allow for content moderation without any legal supervision. As private-sector actors, platforms can put in place Terms of Service allowing for further regulation of content than state laws do, which would effectively give private actors responsibility for defining harmful content[70]. In fact, as pointed out by Keller (2021), this question does not concern solely algorithmic amplification, but instead relates to all issues of content moderation, whether it is hosting or dissemination[71]. **Countering the potential harms of fringe content amplification requires a mix of regulatory and voluntary approaches through collaboration between public and private actors, as well as civil society**[72].

---

[66] See, for instance, the Center for Countering Digital Hate's call to suspend Instagram's recommendation algorithm; discussion of the EU CT Coordinator report's differentiation between hosting and active promotion of content.

[67] Ingram (2021), 'What should we do about the algorithmic amplification of disinformation?'; Keller (2021); Faddoul et al. (2020); RAND Europe interviews.

[68] McNamee (2020); Center for Countering Digital Hate; Barata (2021), 'Moderating Terrorist and Extremist Content'.

[69] RAND Europe interviews; Krasodomski-Jones (2020).

[70] Tech Against Terrorism (2021); Krasodomski-Jones (2020); RAND Europe interviews.

[71] Keller (2021).

[72] RAND Europe interviews.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005619

As this debate is ongoing, existing measures identified by the study team reflect the variety of potential approaches to content moderation and to tackling the malign use of algorithmic amplification. Table 3.1 below provides a non-exhaustive list of measures established by public and private actors, as well as through public-private collaborations. The table presents different types of measures that include, but are not limited to, those stemming from private-sector actors:

- legal or regulatory measures;
- technical measures to limit dissemination of extremist and/or harmful content;
- technical measures to counter the potential impact of extremist and/or harmful content;
- collaborative initiatives bringing together public and private actors; and
- civil society initiatives.

Table 3.1 Overview of measures and practices across the case study Member States

Initiative	Description
Legal or regulatory measures	
Article 3 of The Audiovisual Media Services Directive – EU	This EU directive outlines exceptions to the free transmission of audiovisual media content across Member States, including cases where content contains incitement to hatred based on race, sex, nationality or religion, or otherwise violates human dignity[73].
EU Directive 2017/541 – EU	As part of the intensification of counterterrorism efforts, this 2017 EU directive contains multiple articles that prohibit various terrorist-related offences online. Most notably, Article 5 prohibits the direct or indirect incitement and glorification of terrorism, and Article 21 calls on Member States to 'ensure the prompt removal of online content constituting a public provocation to commit a terrorist offence'[74].
EU Regulation 2021/784 – EU	This EU Regulation requires platforms to remove terrorist content referred by law enforcement authorities from Member States within one hour and introduces obligations for platforms and national authorities to report on the removal of terrorist content, including measures in place, complaints, appeals and penalties. The Regulation will apply as of June 2022.[75]
Proposal – Single Market for Digital Services (Digital Services Act) and	This legislation proposed to the European Parliament aims to increase the scope of existing regulations concerning digital spaces. In particular, the proposal outlines stricter regulation and liability measures for large online platforms. Stakeholders consulted for the proposal expressed

[73] European Parliament (2010), 'Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive)'.

[74] European Commission (2017), 'Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism and replacing Council Framework Decision 2002/475/JHA and amending Council Decision 2005/671/JHA'.

[75] European Parliament and European Council (2021), Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005620

Initiative	Description
amending Directive 2000/31/EC	concern at the potential for algorithmic amplification to promote harmful content and disinformation. They also highlighted the need for algorithmic accountability and transparency, as well as concerns regarding the lack of user empowerment, oversight and enforcement regarding content personalisation through algorithms. There was consensus on the need for effective oversight and accountability mechanisms. The stakeholders did not, however, recommend that harmful yet legal content should be removed[76].
Network Enforcement Act (NetzDG) – DE	This German law came into effect on 1 January 2018. It requires platforms to delete illegal 'hate speech' content. Failure results in fines up to EUR 58 million[77].
Article 8 of the Law 34/2002 on Information Society Services and Electronic Commerce (LSSI) – ES	This law permits Spanish ministries to request the removal or blocking of online content to 'the safeguard of public order, criminal investigation, public security and national defence'. It also grants power of oversight to law enforcement. Articles 510 and 578.4 of the Spanish Criminal Code also authorise the removal of content classified as hate speech or terrorism[78].
Central Office for the Fight against Crime Relating to Information and Communication Technology (OCLCTIC) – FR	The OCLCTIC is mandated by the French data protection agency to request the removal of suspected terrorist content from digital platforms. Its decisions are based on a 2015 law following the declaration of a state of emergency after the 2015 Paris terrorist attacks. The emergency law grants the interior minister the authority to request immediate blocking of content that 'glorifies or incites acts of terrorism'[79].
Online Safety Bill – UK	In May 2021, the UK government put forwards a draft Online Safety Bill introducing a regulatory framework for 'online harms'. The bill regulates both illegal content and harmful content that service providers determine, on 'reasonable grounds', to present a risk of physically or psychologically impacting children or adults, or a risk of 'significant adverse physical or psychological impact' if disseminated[80].
Technical measures to limit the spread and amplification of extremist content	
Content moderation	Social media platforms have, for the most part, adapted their Terms of Service, which now ban the posting of terrorist and violent content, as well as hateful or abusive content (specifically that depicting children). Some platforms, including Twitter, YouTube and Facebook, have also expanded their list of protected groups in their hate speech regulation beyond the protected categories outlined by the UN or the EU[81]. As a result of increased content moderation capacity, millions of posts and

[76] European Commission (2020), 'Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC'.

[77] German Law Archive (2017), 'Network Enforcement Act (Netzdurchsetzunggesetz, NetzDG)'.

[78] Ramaty, N. (2020), 'The Legal Response of Western Democracies to Online Terrorism and Extremism'.

[79] *'LOI N° 2015-1501 Du 20 Novembre 2015 Prorogeant L'application De La Loi N° 55-385 Du 3 Avril 1955 Relative À L'état D'urgence Et Renforçant L'efficacité De Ses Dispositions'* (2015).

[80] Woodhouse (2021), 'Regulating Online Harms. Briefing Paper', House of Commons Library.

[81] Davey et al. (2019); Twitter (2021), 'The Twitter Rules'.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005621

Initiative	Description
	accounts are removed on a monthly basis. For example, between January and March 2021, YouTube removed 2 230 310 accounts, 2% of which contained hateful content or incitements to terrorism or violence[82]. The challenge remains for small and medium-sized platforms to acquire sufficient content moderation capacity to address terrorist or violent content adequately.
EU Code of Conduct	This European Commission initiative sets out standards to counter illegal hate speech online through content removal and platform transparency. As of July 2021, all major social media platforms and the professional platform LinkedIn are participating in the initiative. A 2019 evaluation showed that companies on average reviewed 90% of the notifications within 24 hours and removed 71% of reviewed material[83].
Internet Referral Units (IRUs)	IRUs are law enforcement units that use the flagging function on online platforms to report content violating their Terms of Service. Europol created an EU-wide IRU in July 2015 following the recommendation of the EU CT coordinator[84].
Institute for Strategic Dialogue: Counter Conversations Programme	This is an intervention programme on Facebook that implements practices from deradicalisation programmes in the offline space, to reach radicalised individuals from jihadist and far-right extremist movements on social media platforms. The method entails automatic identification of individuals engaging with extremist content, who receive direct messages from mentors (councillors or former extremists among others) inviting them to have a conversation about extremist beliefs. With a one-in-five response rate, and a sustained conversation rate of between two thirds and three quarters, the pilot programme shows promising results for developing alternative methodologies to content removal to prevent radicalisation in online spaces[85].
Technical measures to counter the potential impact of extremist content online	
Search redirect	The **Redirect Method** is a partnership project established in 2018 by Moonshot, in collaboration with the Anti-Defamation League and the Gen Next Foundation, to target ISIS-related content. Currently active on YouTube, the initiative targets users searching for terrorist content in English and Arabic, redirecting them towards information resources countering the terrorist message, and exit-programmes[86]. Similarly, the **Redirect Initiative**, was adopted by Facebook and Instagram across the US, UK, Indonesia, Australia and Germany to prevent the spread of violent extremism and online radicalisation. Users who search for terms commonly associated with violent extremist or hate groups are redirected towards support programmes,

[82] YouTube (2021), 'Transparency Report'.

[83] European Commission (2016), 'EU Code of Conduct'.

[84] Center for Democracy and Technology (2016), 'Comments of the Center for Democracy & Technology To the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression in the consultation on 'Freedom of expression and the private sector in the digital age'.

[85] Institute for Strategic Dialogue (2018), 'Counter-Conversations: A model for direct engagement with individuals showing signs of radicalisation online'.

[86] Moonshot (2021), 'The Redirect Method'.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005622

Initiative	Description
	such as Life After Hate, with the aim of providing at-risk individuals with resources to facilitate deradicalisation[87].
Peer to Peer Facebook Global Digital Challenge	This campaign was established by Facebook in 2015, in collaboration with EdVenture Partners, to engage university students biannually to create anti-extremism campaigns competitively on social media platforms. The objective of the competition is to create targeted campaigns aimed at young at-risk individuals. The initiative is currently active in 75 countries with a quarter of the campaigns being developed after the programme[88].
Collaborative initiatives	
The Christchurch Call to Action	This international action plan against terrorist use of the internet was initiated by the governments of New Zealand and France in the wake of the 2019 Christchurch attack, and includes social media companies. The objectives of the 2019 collaborative summit include swift takedown of terrorist and violent extremist content; increased transparency by digital platforms in the removal of content; and preventing algorithms from driving extremist ideology online. Key signatories include Germany, France, the UK, the US and the Council of Europe[89].
Global Internet Forum to Counter Terrorism (GIFCT)	The GIFCT is a non-governmental organisation founded by Facebook, YouTube, Twitter and Microsoft in 2017. The initiative aims to prevent the spread of terrorist and violent extremist content online by sharing technical solutions among digital platforms; engage with researchers and stakeholders, including governments; and promote knowledge of radicalisation in online and offline spaces. The GIFCT maps recommendation algorithms and maintains a database of hashtags commonly used by extremist groups. It has also established a Content Incident Protocol (CIP) allowing companies to identify and initiate removal processes for terrorist content across digital platforms[90].
EU Internet Forum	Established in 2015, the EUIF is a collaborative platform working to map the latest developments in terrorist and violent extremist online operations, as well as child sexual abuse on the internet. EU and European Free Trade Association (EFTA) Member States participate alongside Europol, the RAN network, GIFCT, and relevant academic experts and networks, to tackle these threats and prevent future escalation[91].
European Commission's Civil Society Empowerment Programme	This sub-initiative of the EUIF aims to tackle terrorist and violent extremist content by providing avenues for partnership, knowledge-sharing and capacity building among social media platforms and civil society organisations, including grass root campaigners. The programme currently runs training in strategic online campaigning

[87] Counterspeech (2021b), 'The Redirect Initiative'.

[88] Counterspeech (2021a), 'Peer to Peer (P2P): Facebook Global Digital Challenge'.

[89] Christchurch Call to Action (2019).

[90] Global Internet Forum to Counter Terrorism (2021).

[91] European Commission (2015), 'EU Internet Forum: Bringing together governments, Europol and technology companies to counter terrorist content and hate speech online'.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005623

Initiative	Description
	aimed at audiences at high risk of radicalisation[92].
Civil society initiatives	
Knowledge Sharing Platform – Tech Against Terrorism	This is an online resource bank for tech companies providing an overview of content standard regulation, transparency requirements and comprehensive lists of symbols, language and hashtags commonly used by designated terrorist groups. The platform's objective is to increase tech companies' awareness of terrorist use of smaller platforms, and to improve transparency and accountability for removing violent extremist content[93].
Terrorist Content Analytics Platform – Tech Against Terrorism	This is a dataset with a comprehensive overview of terrorist content, allowing tech companies and smaller platforms to detect the presence of extremist organisations and prevent the spread of texts and symbols associated with them. T-CAP has currently alerted 55 companies of over 5 000 malign URLs. Access to the database will be expanded to include academic researchers and civil society organisations, but not governments[94].
Online Civil Courage Initiative	The OCCI is an NGO working across the UK, Germany and France to support broad civil society responses to online hate and extremism. Partnering with Facebook, the OCCI provides skills training and best practice sharing in national languages for counter-speech initiatives, coordinating collaboration between subject experts, activists and online content creators to deliver campaigns against disinformation and terrorism online[95].
AlgoTransparency	This is an initiative founded to address the lack of transparency on the specific functions of algorithmic amplification systems used by social media platforms. AlgoTransparency investigates how platforms use algorithmic amplification and the risks of the practice, as well as advocating for increased regulation of algorithmic functions across the EU[96].

[92] European Commission (2021), 'Civil Society Empowerment Program'.

[93] Tech Against Terrorism (2021b), 'Launch Event: The Knowledge Sharing Platform'.

[94] Tech Against Terrorism (2021c), 'Frequently Asked Questions'.

[95] Institute for Strategic Dialogue (2021b), 'The Online Civil Courage Initiative (OCCI)'.

[96] AlgoTransparency (2021), 'Our Impact'.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005624

OVERVIEW OF GAPS, OPPORTUNITIES AND CHALLENGES IN ADDRESSING THE MALIGN USE OF ALGORITHMIC AMPLIFICATION

This section discusses the gaps and challenges identified with regard to existing and proposed measures for tackling the malign use of algorithmic amplification (Section 4.1), as well as opportunities and recommendations that policymakers should consider adopting (Section 4.2).

CURRENT CHALLENGES

Literature reviewed and consultations carried out by the study team have outlined a number of key challenges (outlined below) in tackling the malign use of algorithmic amplification, as well as related issues of online content moderation.

1. **Challenges linked to a lack of understanding of how online platforms work.** As discussed in Section 2.1, there is a gap in understanding among stakeholders outside the tech communities of algorithmic amplification and other processes that are central to the functioning of online platforms. Interviews conducted by the study team highlighted that this constrains, both at policy and practitioner level, discussions about tackling the malign use of algorithmic amplification of terrorist and extremist content online[97]. Furthermore, **existing research points to a lack transparency about the functioning of major online and social media platforms, including the parameters for algorithm recommendations, as well as algorithms used for content moderation.** This limits attempts to better understand the impact of content disseminated and amplified online. It also makes it difficult to assess the effectiveness of the measures in place[98]. Stark and Stegmann (2020) highlight that researchers are often limited to analysing Twitter data, as other platforms such as Facebook are not as accessible. **Better access to data and processes used by online platforms** could also help in understanding whether and to what extent a distinction can be made between inadvertent amplification of terrorist and violent extremist content by companies and deceitful manipulation of algorithms by terrorists and extremists. It could also further help in investigating coordinated disinformation campaigns and other manipulation attempts by malicious actors. However, it remains to be determined how and to which actors such data could be made available without risking further manipulation[99].

2. **A fragmented legal framework on issues of content moderation and censorship in the EU and globally.** As discussed in Section 3, legal measures to limit the spread of terrorist and violent extremist content, as well as harmful content, differ between EU states and globally. Analyses of existing and potential regulatory models for restricting the dissemination of harmful and illegal content outline these differences, and point to

[97] RAND Europe interviews.

[98] Gorva et al. (2020), 'Algorithmic content moderation: Technical and political challenges in the automation of platform governance'; McNamee, R. (2020); Stark, B. & Stegmann, D. (2020).

[99] Stark, B. & Stegmann, D. (2020).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005625

different limitations and shortfalls within them[100]. For example, while certain countries may be able to regulate how legal but harmful content is distributed, the First Amendment of the US Constitution limits the regulation of speech online, including dissemination methods. However, similar discussions are currently ongoing in the US with the potential to introduce liability for amplifying illegal content[101]. Lack of a homogeneous legal approach, and clear guidelines on how content moderation could be implemented within the remit of legal provisions, could lead to difficulties for platforms to implement such regulations, as well as oversight challenges[102].

3. **Challenges related to issues of freedom of speech.** Beyond the challenges in regulatory differences, the very question of content regulation touches on questions of fundamental rights, such as the right to freedom of opinion and expression, especially when considering content that is not defined as illegal or of an overtly terrorist nature[103]. As outlined in Section 3, **a key challenge lies in the definition of harmful content, and concerns have been raised at the idea of enabling online platforms and private-sector actors to define what harmful content is**[104]. On the one hand, if left for online platforms to decide, this risks transferring a law enforcement responsibility to private actors with no judiciary oversight[105] and exposing platforms to potential accusations of infringing fundamental rights[106]. On the other hand, governments setting specific restrictions on what content is seen as harmful risks setting precedents for potential censorship by government, with particular risks in non-democratic contexts. Concerns have been raised, for example, with regard to a lack of transparency on governments' involvement in the GIFCT and the design of classifications and justification for content takedown [107]. For example, the GIFCT has not made its guidelines for determining what constitutes extremist content public, raising concerns over potential biases and mistakes in content removal. Some stakeholders have expressed concern that journalistic content that covers extremism and terrorism, as well as political satire, could potentially risk removal[108].

RECOMMENDATIONS

Based on challenges identified through the review of existing literature and stakeholder consultations, the following section outlines recommendations proposed for consideration by EU and Member State policymakers.

A - Adopt a systemic approach to tackling online dissemination of extremist content

Research conducted by the study team has highlighted that algorithmic amplification is only one element of a wider issue linked to the dissemination of extremist content online. As outlined under Section 2.2, terrorist and extremist groups have migrated to smaller, more opaque platforms and adopted innovative techniques as a result of more stringent content regulation policies on mainstream social media platform. The interviewees consulted

---

[100] Keller, D. (2021), 'Amplification and its Discontents; Ramaty, N. (2020), 'The Legal Response of Western Democracies to Online Terrorism and Extremism'; RAND Europe interviews.

[101] For further information about ongoing discussions in Congress, see ITIF (2021), 'Podcast: How Section 230 Shapes Content Moderation, With Daphne Keller'.

[102] RAND Europe interviews.

[103] Hadavas, C. (2020), 'The future of free speech online may depend on this database'; De Gregorio (2019), 'Free speech in the age of online content moderation'; UN Counter-Terrorism Committee Executive Directorate (CTED) (2020).

[104] Techdirt (2021).

[105] Barata, J. (2021); Stark, B. & Stegmann, D. (2020); Krasodomski-Jones (2020); Gorwa et al. (2020).

[106] RAND Europe interviews. See, for instance, the case brought against Facebook at the Court of Rome after Facebook suspended the account for Italian neo-fascist party Casa Pound: Columbia University (2020), 'Facebook v. CasaPound'.

[107] Hadavas, C. (2020); Barata, J. (2021).

[108] Hadavas, C. (2020); Barata, J. (2021).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005626

highlighted the need to **develop integrated strategies that require and support smaller platforms to implement measures such as content removal and so-called deplatforming (i.e. the removal of a group, organisation or individual from online platforms and fora used to share their ideas and content)**. These strategies have already been implemented on platforms such as Facebook[109], and this could be done by leveraging existing fora and mechanisms established by the GIFCT, the EUIF or Tech Against Terrorism.

B - Advance understanding of algorithmic amplification and its impacts through knowledge-oriented initiatives.

More broadly, implementing a systemic approach should also entail authorities **developing a more nuanced understanding of the functioning of algorithmic amplification at a technical level, including how algorithmic amplification could be leveraged maliciously across platforms or on individual sites.** The EU and its Member States should therefore leverage existing platforms and fora (e.g. EUIF and GIFCT) to **strengthen collaboration with private-sector actors with a view to enabling research on the technical aspects of this phenomenon, as well as on its impacts at an individual and societal level**[110]. This would require researchers to obtain better access to algorithm data as well as general data on users' behaviour.

Research conducted for this study highlighted a need to develop and strengthen the evidence base on **understanding the potential impacts** of the malign use of algorithmic amplification for terrorist and fringe content, as well as on **the effectiveness of existing measures** to tackle these phenomena. This includes measures such as **content-matching techniques, as well as measures aimed at mitigating the effect of potentially harmful content. One example is search redirect methods**, which promote exit programmes or counter-narratives for specific types of content[111]. There is a need to provide more transparent access to data for researchers, and for the design of experimental or quasi-experimental evaluative approaches to online interventions that tackle such phenomena.

In order to address gaps in the evidence base outlined above (e.g. on the functioning of algorithmic amplification, its technical aspects, impacts at the individual and societal levels, and the effectiveness of existing and proposed measures), it is essential to provide **better access to data for researchers and civil society**. As outlined above, a critical challenge to advancing collective understanding of algorithmic amplification via knowledge-oriented initiatives lies in determining who could access relevant commercial and sensitive data and how without the risk of further manipulation or theft. First steps in this direction could include:

- Establishing requirements for companies to invest in knowledge sharing activities for policymakers and civil society to foster a more widespread understanding of the phenomenon and its impact; and
- Monitoring and conducting impact assessments of existing fora, such as the EUIF and GIFCT, and/or of data access frameworks planned as part of the current Digital Services Act proposal to ensure they contribute to advancing knowledge and contributing to a more transparent understanding of algorithmic amplification and its challenges.

---

[109] RAND Europe interviews; Roger (2020); Davey et al. (2020).

[110] EU Counter-Terrorism Coordinator (2020).

[111] Barata, J. (2021); RAND Europe interviews.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005627

C - Support the implementation of technological solutions through the definition of a framework for content moderation.

As discussed in Section 3, technological solutions exist and can be put in place to limit the dissemination and amplification of terrorist or fringe content. For these measures to be implemented in a concerted, effective and transparent manner, online platforms, including smaller online platforms that are increasingly used by fringe groups, would need to rely on **a defined content moderation framework**. Interviewees consulted in the context of this study highlighted a need for **clear designations and guidelines (which are currently lacking) for platforms to tackle the dissemination of terrorist, extremist and fringe content effectively**[112].

Building on current efforts from Europol and the EUIF, specific elements of such a framework could include: **(1) a clear definition of terrorist and violent extremist content that should be removed; and (2) guidelines on content that should be monitored (e.g. logos, visual identities and names of extremist organisations) to apply soft actions such as deranking or demonetisation**. This could also include requirements and support for platforms to improve their content moderation capabilities.

D - Investigate alternative legal measures to those targeting amplification through independent research.

Research and consultations conducted for this study highlighted the limited evidence base available on the impact of algorithmic amplification. Some interviewees also questioned the validity of developing policies focusing on algorithmic amplification in the absence of this[113]. Several studies have proposed regulatory frameworks moving away from censorship and limits on dissemination. These include **proposals for user-based moderation**, whereby users would be able personalise their recommendation algorithm and/or be involved in the content moderation process by being notified, offered the opportunity to appeal their case and incentivised to participate in content moderation efforts[114].

Alternatively, others recommend tackling this issue through **approaches that are grounded in privacy and consumer protection**, moving away from censorship-based approaches and instead focusing on individual rights[115]. Such alternatives, including their potential implications, challenges and impacts, should be investigated through evidence-based research. That should include pilot programmes designed to evaluate the potential relevance, effectiveness and sustainability of these approaches on a large scale.

E - Further invest in digital skills and capabilities for law enforcement to foster collaboration with platforms, including at national level.

Mechanisms exist for collaboration between online platforms and law enforcement, notably through law enforcement outreach teams among large tech companies, as well as through some Internet Referral Units. However, interviewees from different stakeholder groups consulted in this research noted that a lack of or limited digital skills and capabilities among law enforcement agencies was a factor that could potentially limit collaboration. In particular, interviewees pointed to a limited availability of a critical mass of personnel at law enforcement agencies with knowledge and expertise in data, processes and mechanisms

---

[112] RAND Europe interviews.

[113] RAND Europe interviews.

[114] De Gregorio (2019); Krasofomski-Jones (2020).

[115] Keller, D. (2021); Techdirt (2021); Faddoul et al (2020).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005628

used by platforms for algorithmic amplification[116]. **In addition to research designed to improve understanding of these phenomena among public authorities and law enforcement agencies, there could be merit in investing in education training and exercise interventions aimed at building the digital skills and capabilities of law enforcement personnel**. This could enable better and wider collaboration on these matters between law enforcement and private-sector actors.

[116] RAND Europe interviews. Cedefop (2018). Galai et al. (2019).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005629

ANNEX A. REFERENCES AND INTERVIEWS

This Annex presents an overview of sources consulted by the study team for this research paper (Section A.1), as well details on the interview approach adopted by the study team (Section A.2).

REFERENCES

AlgoTransparency. 2021. 'Our Impact'. Accessed on 14 July 2021: https://www.algotransparency.org/our-impact.html

Ayad, M. 'Islamic State terrorist propaganda is going viral on Facebook.' Wired. Accessed on 30 July 2021: https://www.wired.co.uk/article/islamic-state-terrorism-facebook

Barata, J. 2021. 'Moderating Terrorist and Extremist Content.' Accessed on 16 July 2021: https://www.voxpol.eu/moderating-terrorist-and-extremist-content/

BBC. 2015. 'Fox News explains why it showed Jordan pilot video'. Accessed on 14 July 2021: https://www.bbc.co.uk/news/world-us-canada-31013455

BBC. 2019. 'Morocco hikers: Three get death penalty for Scandinavian tourist murders'. Accessed on 14 July 2021: https://www.bbc.co.uk/news/world-africa-49029505

Bellasio, J., Hofman, J., Ward, A., Nederveen, F., Knack, A., Meranto, A., Hoorens, S. 2018. 'Counterterrorism evaluation: Taking stock and looking ahead'. Accessed on 14 July 2021: https://www.rand.org/pubs/research_reports/RR2628.html

Bellasio, J., Slapakova, L., Quimbre, F., Stockwell, S., Silfversten, E. 2021. 'Human Rights in the Digital Age'. Accessed on 10 September 20201: https://www.rand.org/pubs/research_reports/RRA1152-1.html

Bilton, N. 2019. The Downfall of Alex Jones Shows How the Internet Can Be Saved. *Vanity Fair*. Accessed on 16 July 2021: https://www.vanityfair.com/news/2019/04/the-downfall-of-alex-jones-shows-how-the-internet-can-be-saved

Bipartisan Policy Center, Task Force on Terrorism and Ideology. 2018. 'Digital Counterterrorism: Fighting Jihadists Online'. Accessed on 14 July 2021: https://bipartisanpolicy.org/download/?file=/wp-content/uploads/2019/03/BPC-National-Security-Digital-Counterterrorism.pdf

Bossong, R. 2021. 'The Next Steps for EU Counterterrorism Policy. Evolving Threats of Jihadism, Right-wing Extremism, and Transatlantic Cooperation.' Stiftung Wissenschaft und Politik. Accessed on 16 July 2021: https://www.swp-berlin.org/10.18449/2021C20/

Burwell, F. 2021. 'Free speech and online content: What can the US learn from Europe?' The Atlantic Council. Accessed on 14 July 2021: https://www.atlanticcouncil.org/blogs/new-atlanticist/free-speech-and-online-content-what-can-the-us-learn-from-europe/

Cedefop. 2018. 'Insights into skill shortages and skill mismatch'. Luxembourg: Publications Office of the European Union; No 106. Accessed on 10 September 20201: https://www.cedefop.europa.eu/files/3075_en.pdf

Center for Countering Digital Hate. 2020. 'Malgorithm'. Accessed on 14 July 2021: https://www.counterhate.com/malgorithm

Center for Democracy and Technology. 2016. 'Comments of the Center for Democracy & Technology To the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression in the consultation on 'Freedom of expression and the private sector in the digital age'. Accessed on 14 July 2021:

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_005630

https://www.ohchr.org/Documents/Issues/Expression/PrivateSector/CenterDemocracyTechnology.pdf

Collins, N. 2020. 'Jacinda Ardern calls for 'ethical algorithms' to combat online extremism. What this means'. The Conversation. Accessed on 14 June 2021: https://theconversation.com/jacinda-ardern-calls-for-ethical-algorithms-to-combat-online-extremism-what-this-means-160986

Columbia University, 2020. 'Facebook v. CasaPound. Case Analysis. Accessed on 15 July 2021: https://globalfreedomofexpression.columbia.edu/cases/casapound-v-facebook/

Counterspeech. 2021a. 'Peer to Peer (P2P): Facebook Global Digital Challenge'. Accessed on 14 June 2021: https://counterspeech.fb.com/en/initiatives/p2p-facebook-global/

Cox, K., Ogden, T., Jordan, V., Paille, P. 2021. 'COVID-19, Disinformation and Hateful Extremism'. RAND Europe. Accessed on 16 July 2021: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993842/RAND_Europe_Final_Report_Hateful_Extremism_During_COVID-19_Final.pdf

Crawford, B. 2020. 'Coronavirus and conspiracies: how the far right is exploiting the pandemic'. King's College London. Accessed on 16 July 2021: https://www.kcl.ac.uk/coronavirus-and-conspiracies-how-the-far-right-is-exploiting-the-pandemic

Davey, J., Saltman, E. M., Birdwell, J. in Herman, L., Muldoo, J. *(ed)* 2019. 'The mainstreaming of far-right extremism online and how to counter it. A case study on UK, US and French elections'. in *Trumping the Mainstream.* Routledge Press, London.

De Gregorio, G. 2019. 'Free speech in the age of online content moderation'. Voelkerrechtsblog. Accessed on 14 July 2021: https://voelkerrechtsblog.org/de/free-speech-in-the-age-of-online-content-moderation/

EU Counter-Terrorism Coordinator. 2020. 'The role of algorithmic amplification in promoting violent and extremist content and its dissemination on platforms and social media'. Accessed on 14 July 2021: https://data.consilium.europa.eu/doc/document/ST-12735-2020-INIT/en/pdf

European Commission. 2015. 'EU Internet Forum: Bringing together governments, Europol and technology companies to counter terrorist content and hate speech online'. Accessed on 14 July 2021: https://ec.europa.eu/commission/presscorner/detail/en/IP_15_6243

European Commission. 2017. 'Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism and replacing Council Framework Decision 2002/475/JHA and amending Council Decision 2005/671/JHA'. Accessed on 14 July 2021: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32017L0541

European Commission. 2020. 'Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC'. Accessed on 29 July 2021: https://ec.europa.eu/info/sites/default/files/proposal_for_a_regulation_on_a_single_market_for_digital_services.pdf

European Commission. 2021. 'Civil Society Empowerment Program'. Accessed on 14 July 2021: https://ec.europa.eu/home-affairs/what-we-do/networks/radicalisation_awareness_network/civil-society-empowerment-programme_en

European Parliament. 2010. 'Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive)'. Accessed on 14 July 2021: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32010L0013

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_005631

European Parliament and European Council. 2021. Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online. AS of 1 September 2021 : https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2021.172.01.0079.01.ENG

Facebook. 2021. 'Dangerous Organizations: Terrorism and Organized Hate'. Transparency Centre. Accessed on 30 July 2021: https://transparency.fb.com/data/community-standards-enforcement/dangerous-organizations/facebook#CONTENT_ACTIONED

Faddoul, M., Chaslot, G., Farid, H. 2020. 'A longitudinal analysis of YouTube's promotion of conspiracy videos'. Accessed on 14 July 2021: https://www.semanticscholar.org/paper/A-Longitudinal-Analysis-of-YouTube%27s-Promotion-of-Faddoul-Chaslot/e9b8266ec5d94660dde5f89b4e01586dcddf2515

German Law Archive, 2017. 'Network Enforcement Act (Netzdurchsetzunggesetz, NetzDG)'. Accessed on 14 July 2021: https://germanlawarchive.iuscomp.org/?p=1245

Ghosh, D. and Scott, B. 2018. '#DIGITALDECEIT The Technologies Behind Precision Propaganda on the Internet'. Harvard Kennedy School. Accessed on 14 July 2021: https://www.newamerica.org/pit/policy-papers/digitaldeceit/

Gorwa, R., Binns, R., Katzenbach, C. 2020. 'Algorithmic content moderation: Technical and political challenges in the automation of platform governance.' Accessed on 14 July 2021: https://journals.sagepub.com/doi/full/10.1177/2053951719897945

Galai, K., Retter, L., Muravska, J., Kepe, M., Lynch, A., Knack, K., Bellasio, J., Ward, A., Meranto, S., Maistro, D. 2019. 'Vision on defence related skills for Europe today and tomorrow'. Accessed on 10 September 2021 : https://ec.europa.eu/docsroom/documents/36238

Ganesh, B. and J. Bright. 2019. 'Extreme Digital Speech: Contexts, Responses, and Solutions'. Vox Pol. Accessed on 30 July 2021: https://www.voxpol.eu/download/vox-pol_publication/DCUJ770-VOX-Extreme-Digital-Speech.pdf

Hutchinson, J., Amarasingam, A., Scrivens, R., Ballsun-Stanton, B. 2021. 'Mobilizing extremism online: comparing Australian and Canadian right-wing extremist groups on Facebook'. Accessed on 14 July 2021: https://www.tandfonline.com/doi/full/10.1080/19434472.2021.1903064

Ignatidou, S. 2019. 'AI-Driven Personalization in Digital Media: Political and Societal Implications.' Chatham House. Accessed 10 September 2021: https://www.chathamhouse.org/sites/default/files/021219%20AI-driven%20Personalization%20in%20Digital%20Media%20final%20WEB.pd

Information Technology & Innovation Foundation (ITIF). 2021. 'Podcast: How Section 230 Shapes Content Moderation, With Daphne Keller'. Accessed on 30 July 2021: https://itif.org/podcast-how-section-230-shapes-content-moderation-daphne-keller

Ingram, M. 2021. 'What should we do about the algorithmic amplification of disinformation?'. *Columbia Journalism Review.* Accessed on 14 July 2021: https://www.cjr.org/the_media_today/what-should-we-do-about-the-algorithmic-amplification-of-disinformation.php

Institute for Strategic Dialogue. 2018. 'Counter-Conversations: A model for direct engagement with individuals showing signs of radicalisation online'. Accessed on 14 July 2021: https://www.isdglobal.org/isd-publications/counter-conversations-a-model-for-direct-engagement-with-individuals-showing-signs-of-radicalisation-online/

Institute for Strategic Dialogue. 2021a. 'Recommended Reading: Amazon's algorithms, conspiracy theories and extremist literature'. Accessed on 14 July 2021: https://www.isdglobal.org/isd-publications/recommended-reading-amazons-algorithms-conspiracy-theories-and-extremist-literature/

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_005632

Institute for Strategic Dialogue. 2021b. 'The Online Civil Courage Initiative (OCCI)'. Accessed on 14 July 2021: https://www.isdglobal.org/isd-programmes/online-civil-courage-initiative-occi/

Institute for Strategic Dialogue. 2021c. /Hatescape: An In-Depth Analysis of Extremism and Hate Speech on TikTok'. Accessed on 1 September 2021 : https://www.isdglobal.org/wp-content/uploads/2021/08/HateScape_v5.pdf

Keller, D. 2021. 'Amplification and its Discontents'. Columbia University. Accessed on 14 July 2021: https://knightcolumbia.org/content/amplification-and-its-discontents

Kofman, A., Tseng, F. and Weigel, M. 2020. 'The Hate Store: Amazon's Self-Publishing Arm Is a Haven for White Supremacists'. ProPublica. Accessed 1 September 2021: https://www.propublica.org/article/the-hate-store-amazons-self-publishing-arm-is-a-haven-for-white-supremacists

Krasodomski-Jones, A. 2020. 'Everything in Moderation: Platforms, communities and users in a healthy online environment'. Accessed on 17 July 2021: https://demos.co.uk/project/everything-in-moderation-platforms-communities-and-users-in-a-healthy-online-environment/

Legifrance. 2015. '*LOI N° 2015-1501 Du 20 Novembre 2015 Prorogeant L'application De La Loi N° 55-385 Du 3 Avril 1955 Relative À L'état D'urgence Et Renforçant L'efficacité De Ses Dispositions*'. 2015. Accessed on 14 July 2021: https://www.legifrance.gouv.fr/loda/id/JORFTEXT000031500831/

Lieberman, A. V. 2017. 'Terrorism, the Internet, and Propaganda: A Deadly Combination', *Journal of National Security Law & Policy 9:95*. Accessed on 30 July 2021: https://jnslp.com/wp-content/uploads/2017/04/Terrorism_the_Internet_and_Propaganda_FINAL.pdf

Macklin, G. 2019. 'The Christchurch Attacks: Livestream Terror in the Viral Video Age'. CTC Sentinel. Accessed on 14 July 2021: https://ctc.usma.edu/wp-content/uploads/2019/07/CTC-SENTINEL-062019.pdf

Miller, C. 2018. 'McInnes, Molyneux, and 4chan: Investigating pathways to the alt-right'. Southern Poverty law Center. Accessed on 14 July 2021: https://www.splcenter.org/20180419/mcinnes-molyneux-and-4chan-investigating-pathways-alt-right

Ramaty, N. 2020. 'The Legal Response of Western Democracies to Online Terrorism and Extremism'. Accessed on 14 July 2021: https://www.voxpol.eu/download/vox-pol_publication/The-Legal-Response-of-Western-Democracies-to-Online-Terrorism-and-Extremism.pdf

Ranstorp, M. 2016. The Root Causes of Violent Extremism. RAN Issue Paper. Accessed on 16 July 2021: https://ec.europa.eu/home-affairs/sites/default/files/what-we-do/networks/radicalisation_awareness_network/ran-papers/docs/issue_paper_root-causes_jan2016_en.pdf

Stark, B., Stegmann, D. 2020. 'Are Algorithms a Threat to Democracy? The Rise of Intermediaries: A Challenge for Public Discourse'. Accessed on 14 July 2021: https://algorithmwatch.org/en/wp-content/uploads/2020/05/Governing-Platforms-communications-study-Stark-May-2020-AlgorithmWatch.pdf

Suzor, N., Myears West, S., Quodling, A., York, J. 2020. What Do We Mean When We Talk About Transparency? Toward Meaningful Transparency in Commercial Content Moderation. International Journal of Communication. Accessed on 16 July 2021:

Tech Against Terrorism. 2019. 'ISIS use of smaller platforms and the DWeb to share terrorist content'. Accessed on 14 July 2021: https://www.techagainstterrorism.org/2019/04/29/analysis-isis-use-of-smaller-platforms-and-the-dweb-to-share-terrorist-content-april-2019/

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_005633

Tech Against Terrorism. 2021a. 'Position paper Content personalisation and the online dissemination of terrorist and violent extremist content'. Accessed on 14 July 2021: https://www.techagainstterrorism.org/2021/02/17/position-paper-content-personalisation-and-the-online-dissemination-of-terrorist-and-violent-extremist-content/

Tech Against Terrorism. 2021b. 'Launch Event: The Knowledge Sharing Platform'. Accessed on 14 July 2021: https://www.techagainstterrorism.org/event/launch-event-the-knowledge-sharing-platform/

Tech Against Terrorism. 2021c. 'Frequently Asked Questions'. Accessed on 14 July 2021: https://www.terrorismanalytics.org/faq

Techcrunch 2021. At social media hearing, lawmakers circle algorithm-focused Section 230 reform. Accessed on 16 July 2021:https://techcrunch.com/2021/04/27/section-230-bills-algorithms-congress-hearing/

Techcrunch 2021. At social media hearing, lawmakers circle algorithm-focused Section 230 reform. Accessed on 16 July 2021:

Techdirt. 2021. 'Regulating Amplification Is A Lot Harder Than You Think'. Accessed on 14 July 2021: https://podcasts.apple.com/gb/podcast/regulating-amplification-is-a-lot-harder-than-you-think/id940871872?i=1000526656164

Veilleux-Lepage, Y. 2016. 'Paradigmatic Shifts in Jihadism in Cyberspace: The Emerging Role of Unaffiliated Sympathizers in Islamic State's Social Media Strategy'. Accessed on 14 July 2021: https://jtr.st-andrews.ac.uk/articles/10.15664/jtr.1183/print/

Vitale, H. M., Keagle, J.M. 2014. 'A Time to Tweet, as well as a Time to Kill: ISIS' Project of Power in Iraq and Syria'. Institute for National Strategic Studies. Accessed on 14 July 2021: http://ndupress.ndu.edu/Portals/68/Documents/defensehorizon/DH-77.pdf

Wood, M. A. 2017. 'Antisocial media and algorithmic deviancy amplification: Analysing the id of Facebook's technological unconscious'. Accessed on 14 July 2021: https://journals.sagepub.com/doi/10.1177/1362480616643382

Woodhouse, J. 2021. 'Regulating Online Harms. Briefing Paper.' House of Commons Library. Accessed on 14 July 2021: https://researchbriefings.files.parliament.uk/documents/CBP-8743/CBP-8743.pdf

Global Internet Forum to Counter Terrorism. 2021. Accessed on 14 July 2021: https://gifct.org/about/

Counterspeech. 2021b. 'The Redirect Initiative'. Accessed on 14 June 2021: https://counterspeech.fb.com/en/initiatives/redirect/

European Commission. 2016. 'EU Code of Conduct'. Accessed on 14 July 2021: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en

Christchurch Call to Action. 2019. Accessed on 14 June 2021: https://www.christchurchcall.com/

Moonshot. 2021. 'The Redirect Method'. Accessed on 14 July 2021: https://moonshotteam.com/redirect-method/

Stop Hate For Profit. 2021. Accessed on 14 July 2021: https://www.stophateforprofit.org/

The Trust Project. 2021. Accessed on 14 July 2021: https://thetrustproject.org/

Twitter. 2021. The Twitter Rules. Accessed on 14 July 2021 https://help.twitter.com/en/rules-and-policies/twitter-rules

United Nations Office of Counter-Terrorism (UNOCT). 2021. 'Algorithms and Terrorism: The Malicious Use Of Artificial Intelligence For Terrorist Purposes'. Accessed on 30 July 2021:

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_005634

HYPERLINK ""
https://www.un.org/counterterrorism/sites/www.un.org.counterterrorism/files/malicious-use-of-ai-uncct-unicri-report-hd.pdf

UNDP. 2019. 'YouTube Creators for Change' Accessed on 14 July 2021:. https://www.asia-pacific.undp.org/content/rbap/en/home/programmes-and-initiatives/creators-for-change.html

United Nations Counter-Terrorism Committee Executive Directorate (CTED). 2020. CTED Launches Trends Alert on 'Member States Concerned By The Growing and Increasingly Transnational Threat of Extreme Right-Wing Terrorism'. Accessed on 16 July 2021:

Whittaker, J., S. Looney, A. Reed, F. Votta. 2021. 'Recommender systems and the amplification of extremist content'. *Internet Policy Review, 10(2).* Accessed on 30 July 2021: https://doi.org/10.14763/2021.2.1565

YouTube. 2021. 'Transparency Report'. Accessed on 14 July 2021: https://transparencyreport.google.com/youtube-policy/removals

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005635

A.2 INTERVIEW APPROACH

The study team conducted five interviews as part of this research paper and the study team is grateful to everyone who took part. They included a mix of stakeholders representing different perspectives: two academic experts; one law enforcement representative; one data-analysis expert with previous experience of working for Facebook; and two data-analysis experts from civil society. Where consent has been given, names and/or affiliations are listed in the table below.

Table A.1 provides an overview of the questionnaire and topic guide used during the interviews.

Table A.1: Overview of interviews conducted

Interviewee	Date of interview
Redacted – Center for Monitoring, Analysis and Strategy (CeMAS), Redacted	30 June 2021
National Crime Agency of Slovakia, Head of Extremist Screening Centre.	1 July 2021
Redacted – Global Internet Forum to Counter Terrorism, Redacted	2 July 2021
Academic expert.	6 July 2021
Redacted – Tech Against Terrorism, Redacted Policy and Research. Redacted, Tech Against Terrorism, Senior Research Analyst.	12 July 2021

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005636

The table below provides an overview of the interview questions.

FIGURE A.1. INTERVIEW PROTOCOL

- How would you define algorithmic amplification?
- Are you aware of possible uses of algorithmic amplification to promote terrorist and violent extremist content?
  - > If so, can you give specific examples?
  - > What about algorithmic amplification of fringe content (e.g. conspiracy theories, disinformation, etc.) that is not necessarily illegal or outright extremist but potentially harmful?
  - > Are you aware of extremist groups leveraging this opportunity to promote content that may encourage people to link up with them, but that is not outright extremist or violent in nature?
- Are you aware of evidence of the threat posed by this use of algorithmic amplification as a pathway to radicalisation?
- Are there initiatives that you are aware of that try to address this phenomenon and seem promising or effective from your perspective?
  - > To what extent is there measurable evidence or concrete evidence to confirm their effectiveness?
- What do you think could be done to address this phenomenon, both from the perspective of private and public actors?
  - > What about technical solutions or capabilities? Are there specific gaps to be filled?
  - > What about regulatory, legislative and normative approaches?
  - > What about coordination mechanisms?
  - > Does this require a different approach to content moderation compared to what we have seen so far?
  - > Is there any unique contribution that the EU and its Member States could make?

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_005637

Exhibit 287

VIOLENT EXTREMISM AND TERRORISM ONLINE IN 2021

THE YEAR IN REVIEW

Authored by Redacted, and Redacted Redacted Dublin City University and VOX-Pol, for RAN Policy Support

Radicalisation Awareness Network
RAN Policy Support

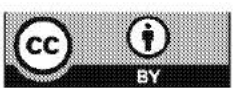
CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011466

LEGAL NOTICE

This document has been prepared for the European Commission however it reflects the views only of the authors, and the European Commission is not liable for any consequence stemming from the reuse of this publication. More information on the European Union is available on the Internet (http://www.europa.eu).

Luxembourg: Publications Office of the European Union, 2021

© European Union, 2021



The reuse policy of European Commission documents is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/).This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

For any use or reproduction of elements that are not owned by the European Union, permission may need to be sought directly from the respective rightholders.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_011467

INTRODUCTION

This report treats developments in the violent extremist and terrorist online scene(s) in the 12-month period from 1 December 2020 to 30 November 2021. It accomplishes this by surveying, describing, and integrating the findings of relevant articles and reports produced by academics, thinktanks, civil society, and governmental organisations; high quality media coverage; and the firsthand experience and primary research of the authors.

The report is divided into three major parts: Part I focuses on the online activities of the extreme right, particularly its European and US variants, Part II on violent forms of jihadism, particularly those linked to the so-called 'Islamic State' (hereafter IS), and Part III details the responses of a range of Internet companies and democratic Western governments to malicious online activity over the course of the last 12 months. We conclude by pointing to a range of issues in the violent extremism and terrorism online realm(s) that bear watching in 2022.

Those familiar with previous VOX-Pol Year in Review reports[1] will know that the structuring of these was slightly different, with analysis of violent jihadi activity preceding discussion of extreme right online activity. The March 2019 Christchurch attacks, a series of subsequent attacks in the United States, Germany, and elsewhere, events online and offline around the US Presidential election, and a general uptick in 'real world' and extreme right online activity globally during the Covid-19 pandemic have all put violent extreme right (online) activity firmly centre stage however, which is reproduced in the structuring of the present report. This does not necessarily mean that the threat from violent jihadism is diminished, but rather reflects the reason behind the increased focus by a range of relevant actors on the risks posed by unfettered online violent extreme right activity.

---

[1] VOX-Pol's *Violent Extremism and Terrorism Online in 2019: The Year in Review* addressed developments in violent extremism and terrorism online from 1 December 2018 to 31 December 2019. It is free-to-access online at https://www.voxpol.eu/download/vox-pol_publication/Violent-Extremism-and-Terrorism-Online-in-2019-The-Year-in-Review.pdf. Our 2018 *Year in Review* is available at https://www.voxpol.eu/download/vox-pol_publication/Year-in-Review-2018.pdf. Our 2017 *Year in Review* is available at https://www.voxpol.eu/download/vox-pol_publication/YiR-2017_Web-Version.pdf and its 2016 equivalent at http://www.voxpol.eu/download/vox-pol_publication/Year-In-Review-WEB.pdf.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011468

PART I.
RIGHT-WING EXTREMIST AND TERRORIST ONLINE CONTENT AND ACTIVITY IN 2021

Concerns about the political fallout of online extreme right activity, including disinformation and radicalisation, began to receive concerted attention from researchers, journalists, policymakers, and others from 2015. The 15 March, 2019 Christchurch terrorist attack mainstreamed these concerns. The New Zealand mosque attack, in which 51 people died, was especially Internet-centric, including a pre-planned online manifesto distribution strategy and Facebook Live video stream. Subsequent events, including a spate of additional right-wing terrorist attacks,[2] the 2020 US Presidential campaign, and the Covid-19 pandemic and associated health and related measures have ensured that the threat posed by contemporary extreme right (online) activity is now under greater scrutiny than ever

Right-wing extremists—like all extremists—structure their beliefs on the basis that the success and survival of the in-group is inseparable from the negative acts of an out-group and, in turn, they are willing to assume both an offensive and defensive stance in the name of the success and survival of the in-group.3 Western right-wing extremism may thus be conceptualised as a racially, ethnically, sexually, and/or gender defined nationalism, which is typically framed in terms of white power and/or white identity (i.e., the in-group) that is grounded in xenophobic and exclusionary understandings of the perceived threats posed by an identified out-group(s), oftentimes a combination of non-whites, Jews, Muslims, immigrants, refugees, members of the LGBTQI+ community,[4] feminists, and/or women.[5] Corrupt political, media, and other elites are often portrayed as championing the out-goup(s) to the detriment of the in-group.

The volume and frequency of production of extreme right online content cannot be measured in the same way as that of jihadis, because the extreme right scene is not dominated by a single group or a discernible number of major franchises or groups, as is the case with violent jihadism. Instead, the extreme right is composed of a shifting and complex overlapping of individuals, groups, movements, political parties, and media organs—both online and traditional—espousing, amongst other views, extreme nationalist, National Socialist/Nazi, fascist, white supremacist, accelerationist, and/or so called 'alt-right' ideology.

Important to acknowledge too is the difficulty of differentiating users, social media accounts, websites, etc., espousing more traditionally violent extremist views (e.g., Nazi or neo-Nazi) from users who hold more radical populist views around, particularly, anti-immigration and Islam, but that are not violent extremists as such. The emergence—online initially—of communities and movements that can have links to the extreme right landscape such as QAnon and incels have only added to this difficulty, as has the increasing

---

[2] Including the April 2019 Poway synagogue attack, the August 2019 El Paso Walmart shooting, and the October 2019 Halle shootings.

[3] J. M. Berger. 2018. *Extremism*. Cambridge, MA: MIT Press.

[4] 'LGBTQI+' is a shorthand reference to describe the lesbian, gay, bi-sexual, transgender, queer, inter-sex, and other sexual identity-based communities.

[5] Barbara Perry and Ryan Scrivens. 2016. 'Uneasy Alliances: A Look at the Right-Wing Extremist Movement in Canada.' *Studies in Conflict & Terrorism* 39(9): 819-841.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011469

overlaps between far- and extreme right activity and agitation around anti-Covid measures (e.g., anti-mask, anti-lockdown, anti-vaccine), both online and off.

Given the volume and diversity of right-wing extremist online content and activity, this section focuses on a selection of right-wing terrorist attacks and other relevant online and offline events, content, and activity that stood out in 2021, but also with reference to 2020, due to the exceptional nature of the global pandemic and its impacts. It begins by reviewing right-wing terrorist attacks and directly adjacent events that occurred in the period under review, with a focus on their online aspects. The next sub-section overviews trends in extreme right online narratives, emphasising the core role of Covid-19, but also briefly addressing QAnon. The third and final sub-section identifies a diversity of platforms and other online spaces and services trafficked by extreme right users in 2021.

RIGHT-WING TERRORISM IN 2021

A serving Belgian soldier, Jürgen Conings, was charged with attempted murder and illegal possession of weapons—thought to be stolen from a military base in Flanders—in a terrorist context after he was connected with threats to kill Belgium's top pandemic virologist, Marc Van Ranst. Conings had been placed on a terrorism watch list in February 2021 as a "potentially violent extremist." He spent from mid-May on the run from Belgian police; his decomposed body, which showed signs of a self-inflicted gunshot wound, was found on 20 June.[6]

Box 1. Memes: A Short Explanation

Memes are pieces of text, images, videos, or some combination of these, oftentimes humorous, which are copied and spread rapidly by Internet users, often with slight variations, the most successful of which enter into popular cultural consciousness. Image macros, **still images upon which a caption has been digitally superimposed**, are the most common form of meme and are widely circulated across social media platforms.

Hateful far- and extreme right memes often feature **distorted or unflattering images of people of colour, Jewish people, and others overlaid with 'humorous' text**. As Siapera *et al.* point out:

In general, **visual elements tend to be recalled faster than audio or text** and retention for images is better and more accurate compared to verbal and textual information. This is important to note here because it implies that **images of hate may be more pernicious than words alone**.[7]

Prior to his death, a Facebook group titled "I love Jürgen Conings" attracted more than 50,000 members. Banned by Facebook, it relocated to the encrypted messaging application Telegram. Posts in the groups praised Conings, including via the use of fascist memes (see Box 1). At least three protests in support of Conings took place between the issuing of a warrant for his arrest and the announcement of his death. One of these coincided with a Brussels protest against Covid-19 measures (see Figure 1).[8]

On 16 March eight people were shot and killed and one wounded in a series of attacks at massage parlours in and around Atlanta, Georgia, USA. Six of the eight killed were Asian women, prompting fears that the attack was

[6] Helen Collis. 'Fugitive Belgian Soldier Jürgen Conings Found Dead.' *Politico*, 20 June, 2021: https://www.politico.eu/article/jurgen-conings-dead-fugitive-belgian-soldier-found/; Evelien Geerts. 'Jürgen Conings: The Case of a Belgian Soldier on the Run shows How the Pandemic Collides with Far-Right Extremism.' *The Conversation*, 16 June, 2021: https://theconversation.com/jurgen-conings-the-case-of-a-belgian-soldier-on-the-run-shows-how-the-pandemic-collides-with-far-right-extremism-162365.

[7] Eugenia Siapera, Elena Moreo, and Jiang Zhou. 2018. *Hate Track: Tracking and Monitoring Racist Speech Online*. Dublin City University: School of Communications and FuJo, p.34: https://www.ihrec.ie/app/uploads/2018/11/HateTrack-Tracking-and-Monitoring-Racist-Hate-Speech-Online.pdf.

[8] *Ibid.*

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011470

racially motivated. However, the attacker, Robert Aaron Long, denied such racial animus once in custody.[9] Rather the suspect informed police that he had a "sexual addiction" and carried out the attacks to remove his "temptation."[10] The charges laid against him included domestic terrorism. While he pleaded guilty to four of the murders and was sentenced to life without parole, he pleaded not guilty to the terrorism and some other charges, which are yet to be adjudicated, but for which the prosecutor is seeking the death penalty.[11]

While some media coverage attempted to tie this attack to hate, incel, or male supremacist movements,[12] whether this can be considered an act of terrorism was unclear at time of writing. In particular, there did not seem to be any online evidence of the perpetrator's involvement in either the right-wing extremist or incel scenes.

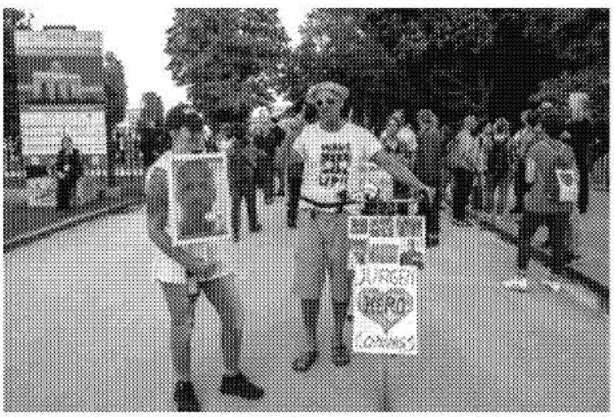


Figure 1. Jürgen Conings' supporters at a protest in Brussels at end of May 2021 (Alamy/Belga News Agency)

On 6 June, a man deliberately rammed a pickup truck into a family of Muslim Pakistani Canadians in London, Ontario, Canada. Of the five people hit, four were killed and the other, a 9-year-old boy, was seriously injured. While the 20-year-old attacker, Nathan Veltman, was reported in the immediate aftermath of the attack as having no known ties to extremist or terrorist groups, the police ruled that the attack was motivated by anti-Muslim hatred and the attacker was charged with murder and terrorism, representing the first time that Canada's anti-terrorism laws were used to prosecute an Islamophobic attack.[13]

While Facebook confirmed to CBC, Canada's national broadcaster, that it took down his profile, Veltman's attack appears thus far to be somewhat of an outlier in that reporters could find no traces of relevant online activity.[14] There is a publication ban in effect on any evidence

---

[9] '8 Dead in Atlanta Spa Shootings, With Fears of Anti-Asian Bias' *The New York Times,* 17 March, 2021: https://www.nytimes.com/live/2021/03/17/us/shooting-atlanta-acworth.

[10] 'Atlanta Spa Shootings: Georgia Man Pleads Guilty.' *BBC News,* 27 July, 2021: https://www.bbc.co.uk/news/world-us-canada-57989425.

[11] Associated Press. 'The Atlanta Spa-Shooting Suspect Pleads Not Guilty In 4 Killings.' *NPR*, 28 September, 2021: https://www.npr.org/2021/09/28/1041137210/atlanta-spa-shooting-suspect-pleads-not-guilty-robert-aaron-long.

[12] '8 Dead in Atlanta Spa Shootings.' For more, on the intersections between hate, inceldom, and male supremacy, see Megan Kelly, Alex DiBranco, and Julia R. DeCook. 2021. *Misogynist Incels and Male Supremacism: Overview and Recommendations for Addressing the Threat of Male Supremacist Violence* (Washington DC: New America): https://www.newamerica.org/political-reform/reports/misogynist-incels-and-male-supremacism/.

[13] Jack L. Rozdilsky. 'The Terrorism Charge Filed in the London Attack is the First of Its Kind in Canada.' *The Conversation*, 17 June, 2021: https://theconversation.com/the-terrorism-charge-filed-in-the-london-attack-is-the-first-of-its-kind-in-canada-162739.

[14] Kate Dubinski and Mark Gollom. 'What We Know About the Accused in the Fatal Attack on a Muslim Family in London, Ont.' *CBC News*, 9 June, 2021: https://www.cbc.ca/news/canada/london/london-murder-suspect-muslim-family-1.6057164.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011471

presented to the courts in this case however, so there is a possibility that some of that evidence may have been sourced online and therefore be revealed when the ban is lifted.

An attack having both extreme right and Covid-19 connections was the murder of a 20-year-old student in Idar-Oberstein, Germany on 18 September, 2021. The student employee of a petrol station was shot dead by a 49-year-old man who told police that he acted "out of anger" having been denied service for refusing to wear a mask while trying to buy beer. The alleged perpetrator, identified only as Mario N, and who turned himself in the morning after the shooting, told police "that he rejected the measures against the coronavirus."[15]

Analysis of Mario N's social media activity, in which he also expressed support for the German far-right party Alternative für Deutschland (AfD) and perpetuated climate denialism, corroborated this. Similar to events in Belgium, the killing was praised in online groups and channels frequented by far-right users and conspiracy theorists, including Germans self-identified as Querdenker or 'Lateral Thinkers.' "One less parasite," read a post noting the victim was a student, while another described the attack as a natural step in the fight against the "Merkel dictatorship," the website of the German daily newspaper Der Tagesspiegel reported.[16]

At issuance again in 2021 was whether incel violence can accurately be described as terrorism.[17] This discussion arose in the context of Jake Davison shooting dead five people, including his mother and a three-year-old girl, in Plymouth, England on 12 August. The shooting spree ended when Davison turned the gun on himself. This attack differs from the Georgia shootings in that Davison had a history of contributing to incel forums and posting YouTube videos in which he appeared to subscribe to incel ideology, including describing himself as "blackpilled" (see Box 2), but in which he denied being an incel. Initially, UK counterterrorism dismissed classifying the shootings as terrorism,[18] but local police later stated that they might be reclassified as such.[19]

---

[15] 'German Cashier Shooting Linked to Covid-19 Conspiracies.' *BBC News*, 22 September, 2021: https://www.bbc.com/news/world-europe-58635103; John Silk. 'Face Mask Killing Sparks Radicalization Fears in Germany.' *DW.com*, 21 September, 2021: https://www.dw.com/en/face-mask-killing-sparks-radicalization-fears-in-germany/a-59252877.

[16] *Ibid*. See also, ISD Germany. 2021. *Wie Instagram scheitert, gegen Hetze vorzugehen: Der Fall Idar-Oberstein* (Berlin: ISD): https://www.isdglobal.org/wp-content/uploads/2021/11/ISD_kurzreport_Instagram-Idar-Oberstein-112021.pdf.

[17] Maura Conway. 2018. *Violent Extremism and Terrorism Online in 2018: The Year in Review* (Dublin: VOX-Pol): https://www.voxpol.eu/download/vox-pol_publication/Year-in-Review-2018.pdf.

[18] Lizzie Dearden. 'Plymouth Shootings Were Not Terror Attack, Senior Police Officer Says.' *The Independent* (UK), 28 September, 2021: https://www.independent.co.uk/news/uk/crime/plymouth-shootings-incel-terror-attack-b1928366.html.

[19] Sky News. 'Plymouth Shootings: Attack Could be Reclassified as Terrorism Over Jake Davison's "Incel" Links.' *Sky News*, 17 August, 2021: https://news.sky.com/story/plymouth-shootings-attack-could-be-reclassified-as-terrorism-over-jake-davisons-incel-links-12383353. Canada has the highest number of recorded incel attacks; in 2019, Canada's Security Intelligence Service identified inceldom as a form of violent misogynistic ideological extremism.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011472

Box 2. Excerpt from Brace's 'A Short Introduction to The Involuntary Celibate Sub-Culture'[20]

Research has shown that the categorical structure of the incel worldview is a **rigid three-tier, immutable, social hierarchy exclusively based on physical appearance**. Here, a minority of **alpha males (Chads)** and **females (Stacys)** are at the top; a majority of **average-looking betas (Normies)** in the middle; and the **exclusively male and minority group of incels at the bottom**.[21]

In this hierarchy, the out-groups (featuring women as well as alphas and betas) are depicted in an extremely negative way and **dehumanised** through the use of negative adjectives and specific terms, such as "femoids" or "roasties," in the case of women. Women are also portrayed as only capable of **simple emotions** (chiefly sexual desire) and guided by **anti-social values**, i.e. cheating on their partners or manipulating men for sex or money.[22] Ironically, while incels view themselves at the bottom of their hierarchy, they **consider all men, including themselves, as superior to women**.

This leads to a central tenet of the incel worldview; the notion of **"pilling"**. This is a concept that also features in some elements of far-right ideology and is borrowed from the 1999 film, **The Matrix**. The idea is that an individual can either take the "blue pill" and remain detached from reality and live life in ignorance, or they can take the **"red pill"** and "wake up" to the true nature of the world and accept that social structures like those above exist. Incels have added a third option to this, the **"black pill"**, which builds on the red pill by adding a nihilistic element to it in stating that these social hierarchies are immutable and that nothing can be done about them.

In the aftermath of the attack, Jonathan Hall, the UK's official Independent Reviewer of Terrorism Legislation, said inceldom "fits rather uneasily into the way the authorities understand ideologies. It seems part of right-wing terrorism but it is not really. In fact, it is quite separate from it. It is a different sort of ideology."[23] Others disagree.[24] While, as Brace has pointed out, "the exact nature of this relationship is yet to be examined empirically," they include, at a minimum, some ideological overlaps, particularly as regards attackers subscribing to both ideologies regularly expressing hate toward women and minorities and both communities exhibiting similar "online sensibilities," including in their use of memes (see Box 1), notion of 'pilling' (see Box 2), etc.[25]

Brace has therefore advised:

[G]iven the evidence currently available, it is perhaps best to view the contemporary far-right not as one coherent ideology but one that is made up of several different "flavours" that share different aspects in their worldview. Perhaps the best way to describe this is as a series of overlapping Venn diagrams, whereby each set is a specific category of the far-right, i.e. white nationalist, neo-Nazi,

---

[20] Lewys Brace. 'A Short Introduction to The Involuntary Celibate Sub-Culture.' CREST, 26 August 2021: https://crestresearch.ac.uk/resources/a-short-introduction-to-the-involuntary-celibate-sub-culture/.

[21] Stephane Baele, Lewys Brace, and Travis Coan. 2019. 'From "Incel" to "Saint": Analyzing the Violent Worldview Behind the 2018 Toronto Attack.' *Terrorism and Political Violence* [Online First].

[22] *Ibid*.

[23] Lauren Kent and Hannah Ritchie. 'Plymouth Shooter Made Misogynist Remarks Echoing the 'Incel' Ideology.' *CNN*, 15 August, 2021: https://edition.cnn.com/2021/08/14/uk/plymouth-shooting-incel-jake-davison-profile-intl/index.html.

[24] Charlotte Littlewood. 'Incel Violence: Should It Be Considered Terrorism?' *European Eye on Radicalization*, 25 October, 2021: https://eeradicalization.com/incel-violence-should-it-be-considered-terrorism/.

[25] Brace, 'A Short Introduction to the Involuntary Celibate Sub-Culture.'

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

militia groups, alt-right, etc. Each set features different aspects of far-right ideology, but not all of them. The incel ideology is one of these sets. [26]

Overall, while there were relatively few unambiguously terrorist extreme right attacks in 2021, there were a number of extreme right-oriented attacks and other events that were considered by researchers, law enforcement, and/or others to fit common definitions of terrorism. Most of the described attacks and other relevant incidents had significant online components, but a not insignificant number (e.g., Atlanta, Ontario) did not appear to.

A series of events that had outsized Internet aspects, including especially in terms of the role of the Internet in identifying those charged and platforms' takedown activity in its aftermath, were those that took place on the US's Capitol Hill on 6 January, 2021. Brace's above-described approach to thinking about the nature of the contemporary extreme right is borne out with respect to these events; the question of what constitutes terrorism also arose in relation to them. For more on whether the events of 6 January constituted terrorism, see Box 3; the roles of the Internet in the Capitol Hill incident are addressed in more detail in the next sub-section.

Box 3. Were the 6 January Capitol Hill Events Terrorism?

A question that arose in the immediate aftermath of 6 January was whether these events could be described as terrorism. Despite the existence of a statutory US definition of domestic terrorism,[27] there is no federally chargeable offence of same. Terrorism charges were never on the table therefore, instead those arrested have been charged with everything from misdemeanors or so-called 'petty offences' to felonies (e.g., "corruptly obstructing, influencing, or impeding an official proceeding," using a dangerous weapon, conspiracy).[28] Journalists, researchers, and others nevertheless posed the question of whether the 6 January events met standard definitions of terrorism. Many agreed that some of the activity met the threshold.[29]

Importantly, two of those who used the term 'terrorism' to describe some of the events at the US Capitol on 6 January were the US President and the Director of the FBI. On 7 January, then US President-elect Biden stated "they weren't protesters. Don't dare call them protesters. They were a riotous mob, insurrectionists, domestic terrorists. It's that basic. It's that simple."[30] Former US President Trump-appointed FBI Director Christopher Wray followed-up by saying in congressional testimony on 2 March that "the violence and destruction that we saw that day...that attack, that siege, was criminal behaviour...and it was behaviour that we, the FBI, view as domestic terrorism."[31]

The Role of Online Pre, During, and Post the 6 January Capitol Hill Events

Five months prior to the events of 6 January, 2021, at the US Capitol, somewhat similar events took place at the German Reichstag. On Saturday, 29 August, 2020 a demonstration

---

[26] *Ibid*.
[27] See section 18 *US Code* § 2331: https://www.law.cornell.edu/uscode/text/18/2331.
[28] Roger Parloff. 'What Do—and Will—the Criminal Prosecutions of the Jan. 6 Capitol Rioters Tell Us?' *Lawfare*, 4 November, 2021: https://www.lawfareblog.com/what-do%E2%80%94and-will%E2%80%94-criminal-prosecutions-jan-6-capitol-rioters-tell-us.
[29] Tanya Mehra and Joana Cook. 'An Attack on the Capitol and Democracy: An Act of Terrorism?' *ICCT Blog*, 11 January, 2021: https://icct.nl/publication/an-attack-on-the-capitol-and-democracy-an-act-of-terrorism/.
[30] Lauren Gambino. 'Biden Decries Trump Mob.' *The Guardian*. 7 January, 2021: https://www.theguardian.com/us-news/2021/jan/07/joe-biden-trump-mob-domestic-terrorists.
[31] Alexander Mallin and Luke Barr. 'FBI Director Says Capitol Assault "Domestic Terrorism," No Evidence of Antifa' (with video). *ABC News*, 2 March, 2021: https://abcnews.go.com/Politics/fbi-director-testifies-time-capitol-assault/story?id=76187365.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011474

attracting an estimated 38,000 attendees, including right-wing extremists, against Germany's coronavirus restrictions took place in Berlin. In the evening, a hundreds-strong section of the crowd stormed the German parliament building, with the protesters initially held off from entering by just three police officers. Ahead of the August 2020 demonstration, a number of public Telegram groups had reportedly called for a "storm on Berlin."[32] Some users had allegedly even posted photos of themselves with weapons.[33] In the event, the German police guarding the Reichstag were quickly reinforced and the building was not breached, the US Capitol Police had a different experience, however.

Washington DC's Capitol Hill is a historic residential neighbourhood in Washington DC that is home to the US Congress. On 6 January, 2021 attendees at a 'Save America' protest against false claims of election irregularities were addressed by outgoing US President Trump who stated "[i]f you don't fight like hell, you're not going to have a country anymore."[34] He encouraged the protesters to march on the US Capitol where a joint session of Congress was taking place to formalise US President-elect Joe Biden's victory.

An unprecedented incident subsequently took place: a section of those present violently and unlawfully entered the US Capitol with a range of weapons, running riot throughout the building, damaging property, looting, and occupying the Senate floor. In front of the Capitol, a mock gallows was constructed and there were calls for the hanging of then-US Vice President Mike Pence, who was overseeing the count inside.[35] The 6 January events resulted in the deaths of four protesters, three from natural causes and a female QAnon supporter who was shot and killed by police as she tried to break into the House chamber. One police officer collapsed and died from a stroke after engaging with the rioters, dozens of others were injured.[36]

The organisation of these events, the events themselves, and post-6 January responses all had significant online components. While this incident took place in the US, numerous aspects of it have resonance in an EU—and wider—context. These include online's role in the run-up to the events; the live streaming of them by participants as they unfolded, including monetisation of these; the key role of especially social media in bringing charges against many of those involved in the events; and the increased takedown activity by major and a panoply of smaller online players post-6 January.

On 19 December, 2020, outgoing US President Donald Trump tweeted a call for his followers to protest in Washington DC on 6 January, the day Electoral College votes were due to be certified by the US Congress and Joe Biden certified as the next US President. "Statistically

---

[32] Katrin Bennhold. 'Far-Right Germans Try to Storm Reichstag as Virus Protests Escalate.' *The New York Times*, 31 August, 2020: https://www.nytimes.com/2020/08/31/world/europe/reichstag-germany-neonazi-coronavirus.html.

[33] *Ibid.*

[34] Julia Jacobo. 'This is What Trump Told Supporters Before Many Stormed Capitol Hill' (with video). *ABC News*, 7 January, 2021: https://abcnews.go.com/Politics/trump-told-supporters-stormed-capitol-hill/story?id=75110558.

[35] Mary Clare Jalonick and Lisa Mascaro. 'GOP Blocks Capitol Riot Probe, Displaying Loyalty to Trump.' *Associated Press*, 28 May, 2021: https://apnews.com/article/michael-pence-donald-trump-capitol-siege-government-and-politics-4798a8617bacf27bbb576a4b805b85d9.

[36] Kenya Evelyn. 'Capitol Attack: The Five People Who Died.' *The Guardian*, 8 January, 2021: https://www.theguardian.com/us-news/2021/jan/08/capitol-attack-police-officer-five-deaths.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

impossible to have lost the 2020 Election," Trump tweeted; "[b]ig protest in DC on January 6th. Be there, will be wild!" (see Figure 2).

Figure 2 - Then Outgoing-US President Trump's 19 December, 2020 Tweet Encouraging Followers to Attend 6 January, 2021 Protest

On Tuesday and Wednesday, 5 and 6 January, protesters arrived in DC in the thousands; they:

> [I]ncluded groups across a spectrum of radicalization: hyperpartisan pro-Trump activists and media outlets; the neo-fascist Proud Boys, a group with chapters committed to racism and the promotion of street violence; unlawful militias from around the country with a high degree of command and control, including the so-called Three Percenters movement; adherents to the collective delusion of QAnon; individuals identifying with the Boogaloo Bois, a loosely organized anti-government group that has called for a second civil war; and ideological fellow travelers of the far-right, who wanted to witness something they believed would be spectacular. Each group overlapped but maintained distinct engagement—and coordination—online.[37]

This online engagement and coordination included spreading support for the protest widely online, including via social media, messaging applications, email lists, and websites. Far-right figures in the US Republican party, other far-right figures, and prominent extreme right users threw their weight behind the protest. These included Rep. Paul Gosar (R-Arizona), StoptheSteal.us organiser Ali Alexander, Women for America First co-founder Kylie Jane Kremer, 8chan's Ron Watkins, and Proud Boy's leader Enrique Tarrio.[38] The online coordination included organisation of travel to DC, including bus and car caravans from across the US.[39]

Many of those present used social media to advertise their presence at the initial 6 January rally and, crucially, events at the Capitol later in the day. This included posting text, photos, and videos across a diversity of social media platforms and messaging applications, in both public and private groups and channels. Of particular note was the relay of events in real time by some of those involved via livestreaming. In some cases, the streams were monetised through on-platform tipping, in others, streamers encouraged viewers to donate GoFundMe, Patreon, and PayPal. For example, prominent white nationalist Tim Gionet ('BakedAlaska') is

---

[37] DFR Lab. 2021. '#StopTheSteal: Timeline of Social Media and Extremist Activities Leading to 1/6 Insurrection.' *JustSecurity*, 10 February, 2021: https://www.justsecurity.org/74622/stopthesteal-timeline-of-social-media-and-extremist-activities-leading-to-1-6-insurrection/.

[38] See *Ibid*. for screenshots of a selection of relevant posts and emails by these and others.

[39] *Ibid*.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011476

estimated to have made over US$2,000 streaming from the Capitol via 'tips' left in DLive's on-platform 'lemons' that can be converted into real money.[40]

The events of 6 January resulted in a major wave of takedowns by social media and other Internet companies during, immediately after, and in the months post-6 January. Facebook banned by then formally outgoing US President Trump's account for 24 hours on 6 January. Facebook followed-up on 7 January with an indefinite ban of a minimum of two weeks.[41] (For more on this and other relevant decisions made by the Facebook Oversight Board in 2021, see sub-section below). On 8 January, 2021 Twitter announced that "[a]fter close review of recent Tweets from the @realDonaldTrump account and the context around them—specifically how they are being received and interpreted on and off Twitter" that it had decided to permanently suspend the account "due to the risk of further incitement of violence."[42] They followed-up by taking extensive enforcement action against QAnon, as well as other groups and movements shown to have played a role (e.g., Oath Keepers). For example, Facebook reported removal of accounts associated with some 600 "militarized social movements."[43]

At least partially taking their cue from these, a range of other platforms took action against Donald Trump, the Trump campaign and associates, and a range of other actors involved in the 6 January events. These included Snapchat, TikTok, and Twitch.[44] It was not just social media, messaging applications, and adjacent platforms that took action post-6 January, however. Stripe, a major payment processor, announced that it would no longer process payments for the Trump campaign as a result of the riots.[45] Airbnb also released a statement that they would not allow their site to be used to organise accommodation for those participating in the protests expected to accompany the new US President Joe Biden's inauguration on 20 January, 2021. They also promised to permanently ban from their platform those who engaged in criminal activity during the 6 January Capitol Hill incident.[46]

---

[40] Kellen Browning and Taylor Lorenz. 'Pro-Trump Mob Livestreamed Its Rampage, and Made Money Doing It.' *The New York Times*, 8 January, 2021: https://www.nytimes.com/2021/01/08/technology/dlive-capitol-mob.html.

[41] Guy Rosen and Monika Bickert. 'Our Response to the Violence in Washington January 6, 2021.' *Facebook Newsroom*, 7 January, 2021: https://about.fb.com/news/2021/01/responding-to-the-violence-in-washington-dc/.

[42] Twitter Inc. 'Permanent Suspension of @realDonaldTrump.' *Twitter Blog*, 8 January, 2021: https://blog.twitter.com/en_us/topics/company/2020/suspension.

[43] Rosen and Bickert, 'Our Response to the Violence in Washington.'

[44] Brian Heater and Taylor Hatmaker. 'Twitch Disables Trump's Channel Until the End of His Term to 'Minimize Harm' During Transition.' *Tech Crunch*, 7 January, 2021: https://techcrunch.com/2021/01/07/twitch-disables-trumps-channel-over-incendiary-rhetoric/;
Salvador Rodriguez. 'Snapchat Will Terminate Trump's Account on Jan.20.' *CNBC*, 13 January, 2021: https://www.cnbc.com/2021/01/13/snapchat-will-ban-trump-on-jan-20.html; Sarah Perez. 'TikTok Bans Videos of Trump Inciting Mob, Blocks #stormthecapitol and Other Hashtags.' *Tech Crunch*, 7 January, 2021: https://techcrunch.com/2021/01/07/tiktok-bans-videos-of-trump-inciting-mob-blocks-stormthecapital-and-other-hashtags/.

[45] AnnaMaria Andriotis, Peter Rudegeair, and Emily Glazer. 'Stripe Stops Processing Payments for Trump Campaign Website.' *The Wall Street Journal*, 10 January, 2021: https://www.wsj.com/articles/stripe-stops-processing-payments-for-trump-campaign-website-11610319116.

[46] 'Airbnb Announces "Capitol Safety Plan" for the Inauguration.' *Airbnb News*, 11 January, 2021: https://news.airbnb.com/airbnb-announces-capitol-safety-plan-for-the-inauguration/.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

An Internet company that suffered a serious setback arising from the events of 6 January was Parler, which lost approximately half of its weekly traffic between July 2020 and July 2021.[47] The explanation for this is almost certainly Parler being banned from Apple's and Google's app stores and suspended by Amazon, their web services provider, in the aftermath of the Capitol's storming. This was due to what the companies described as Parler's inadequate handling of violent threats and hateful content encouraging the riot circulated on its platform. Parler's anxiety to be reinstated in the App Store, where it was the No.1 app when it was taken down on 9 January, shows the sway Apple and other Internet companies have over commercial users of their platforms.[48] While social media companies get the most attention regarding their content moderation practices, companies—sometimes little publicly known—further down the so-called 'tech stack' have powerful roles in decisions around online speech, which are oftentimes overlooked.[49]

Parler was readmitted by Apple to their App Store in mid-May 2017, but only after a prolonged series of negotiation with the latter, knockback of their initial request for readmission in March, and deployment of a new AI-powered moderation system. Readmission was based on an undertaking by Parler that posts labeled "hate" by its new moderation system won't be visible on iPhones or iPads. (Users who view Parler on other smartphones or on the Web will still be able to see posts marked as "hate" by clicking through to them). Parler remains banned by Google's Play Store—though users can still side-load the app on Android via third party providers, including via Parler's own site—and is presently suing Amazon for the denial of web services and reinstatement, but is widely thought unlikely to be successful in the courts.[50]

On the other hand, a company that benefited directly from the 6 January events was Telegram. Telegram received what its founder, Pavel Durov, called "maybe the largest digital migration in human history" in January 2021,[51] becoming the most downloaded mobile app in the world that month.[52] Contributing to this was an influx of far- and extreme right users. The leader of the Proud Boys—which has since been proscribed as a terrorist organisation by Canada—Enrique Tarrio, sang Telegram's praises on his Telegram channel: "Welcome, newcomers, to the darkest part of the web. You can be banned for spamming and porn. Everything else is fair game."[53]

---

[47] Drew Harwell. 'Rumble, a YouTube Rival Popular with Conservatives, Will Pay Creators Who "Challenge the Status Quo."' *Washington Post*, 12 August, 2021: https://www.washingtonpost.com/technology/2021/08/12/rumble-video-gabbard-greenwald/.

[48] Kim Lyons. 'Parler Returns to Apple App Store with Some Content Excluded.' *The Verge*, 17 May, 2021: https://www.theverge.com/2021/5/17/22440005/parler-apple-app-store-return-amazon-google-capitol; Kevin Randall. 'Social App Parler is Cracking Down on Hate Speech — But Only on iPhones.' *The Washington Post*, 17 May, 2021: https://www.washingtonpost.com/technology/2021/05/17/parler-apple-app-store/.

[49] Joan Donovan. 'Navigating the Tech Stack: When, Where and How Should We Moderate Content?' *CIGI*, 28 October, 2019: https://www.cigionline.org/articles/navigating-tech-stack-when-where-and-how-should-we-moderate-content/.

[50] Lyons, 'Parler Returns to Apple App Store'; Randall, 'Social App Parler is Cracking Down on Hate Speech.'

[51] In 14 January, 2021 post by Pavel Durov to his personal Telegram channel: https://t.me/s/durov?q=%E2%80%9Cthe+largest+digital+migration+in+human+history%E2%80%9D.

[52] In 8 February, 2021 post by Pavel Durov to his personal Telegram channel: https://t.me/durov/152.

[53] Michael Schwirtz. 'Telegram, Pro-Democracy Tool, Struggles over New Fans from Far Right.' *The New York Times*, 26 January, 2021: https://www.nytimes.com/2021/01/26/world/europe/telegram-app-far-right.html.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

In the wake of this mass migration, Telegram announced they had blocked hundreds of extreme right user posts calling for violence ahead of the US Presidential inauguration on 20 January.[54] At around the same time, they also banned at least two dozen neo-Nazi channels that had been operating on the platform for some time, most of which had between 2,000 and 10,000 followers.[55] At the end of 2021, Telegram nevertheless remained a preferred app of extreme right users.

An ironic twist was that not only did those who stormed the Capitol on that January Wednesday widely utilise the Internet in the run-up to and during the incident, but much of this same content led to their eventual arrests and prosecutions. According to data collection and analysis undertaken by the Program on Extremism at George Washington University, 82% of the criminal complaints thus far arising from the events of 6 January contained evidence from social media, with 52% stemming from individuals' own social media content.[56] This is because many of the rioters posted a range of content, including their intent to engage in criminal behaviour and/or footage of their actual criminal behaviour as it was happening, sometimes even naming themselves. Thousands of amateur online sleuths archived and analysed this online content and shared their 'leads' with the FBI who acted on some of it.[57] Facebook was reported in the Program on Extremism's research as the most cited platform in the evidence marshalled to charge and prosecute those arrested on 6 January, but YouTube, Instagram, and Parler were also cited.[58]

TRENDS IN EXTREME RIGHT ONLINE ACTIVITY AND NARRATIVES IN 2021

Trends in online extreme right narratives in 2021 must be discussed in the context of also 2020, specifically the COVID-19 pandemic. From the outset of widespread so-called 'lockdowns,' a variety of actors pointed to these as having the potential to increase online radicalisation. Warnings of this type were delivered by the UN Secretary General, António Guterres, in July;[59] then EU Counterterrorism Coordinator, Gilles De Kerchove, in

---

[54] *Ibid*.

[55] Ali Breland. 'Telegram Finally Takes Down Neo-Nazi Channels.' *Mother Jones*, 13 January, 2021: https://www.motherjones.com/politics/2021/01/telegram-nazi-ban/.

[56] Jonathan Lewis and Bennett Clifford. 'Footprints: Social media evidence form the US Capitol siege perpetrators.' *GNET*, 25 January, 2021: https://gnet-research.org/2021/01/25/take-nothing-but-pictures-leave-nothing-but-digital-footprints-social-media-evidence-from-the-us-capitol-siege-perpetrators/.

[57] David Yaffe-Bellany. 'The Sedition Hunters: Amateur Internet Sleuths Have Turned the Washington, DC, Insurrection on Jan. 6 Into the Ultimate Online Manhunt.' *Bloomberg*, 7 June, 2021: https://www.bloomberg.com/features/2021-capitol-riot-sedition-hunters/. See also, Aymann Ismail. 'The Cigarette Smoking Man: I Filmed Inside the Capitol Riot. When I Heard from the FBI, I Faced a Vexing Dilemma.' *Slate*, 12 August 2021: https://slate.com/news-and-politics/2021/08/capitol-riot-footage-photos-online-detectives-fbi.html.

[58] *Ibid*.

[59] United Nations. 'Secretary-General Warns Terrorists Are Exploiting COVID-19 Pandemic, Calls for Vigilance, at Virtual Counter-Terrorism Week Opening.' *Press Release*, 6 July, 2020: https://www.un.org/press/en/2020/sgsm20161.doc.htm.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011479

August;[60] and a variety of police and security agencies over the course of the year.[61] Such cautions continued into 2021, communicated by, among others, German Interior Minister, Horst Seehofer, in June;[62] European Commissioner for Home Affairs, Ylva Johansson, in June;[63] Australia's Home Affairs Minister, Karen Andrews, in September;[64] the UK Security Minister, Damian Hinds, in November;[65] and, again, a variety of police and security agencies.[66]

In 2021, as in 2020, concerns were that anti-Covid measures, especially 'lockdowns,' would lead to people spending more time online and that this would lead to increased numbers of users being exposed to extremist content and narratives, some of whom would probably be radicalised by the latter and their online networking around it. A particular concern was that extreme right groups and users would capitalise on this to reach larger and more varied audiences than previously.

Research confirmed that the pandemic coincided with an increased interest in extremist online content. In the week of 30 March 2020, Moonshot found an average increase in engagement with white supremacist online content of 21 percent in US states with 'stay-at-home orders' of 10 days or more compared with the eight previous months.[67] In April 2020, ISD revealed that mentions of 'corona-chan,' a slang term for COVID-19 then popular with far-right users, had increased massively across both mainstream (e.g., in March, Facebook interactions with the term increased by 1,920%) and fringe (e.g., between February and March, the term was used over 13,000 times on 4chan) social media platforms.[68]

---

[60] Raffaello Pantucci. 2020. 'A View from the CT Foxhole: Gilles de Kerchove, European Union (EU) Counter-Terrorism Coordinator.' *CTC Sentinel* 13(8): https://ctc.usma.edu/a-view-from-the-ct-foxhole-gilles-de-kerchove-european-union-eu-counter-terrorism-coordinator/.

[61] See, for example, National Security and Intelligence Committee of Parliamentarians. 2021. *Annual Report 2020* (Ottawa Her Majesty the Queen in Right of Canada): https://nsicop-cpsnr.ca/reports/rp-2021-04-12-ar/annual_report_2020_public_en.pdf.

[62] Ben Knight. 'Pandemic Spurred Extremism, Says German Domestic Intelligence' *Deutsche Welle*, 15 June, 2021: https://www.dw.com/en/pandemic-spurred-extremism-says-german-domestic-intelligence/a-57906728.

[63] Europol. 'Terrorists Attempted to Take Advantage of the Pandemic, Says Europol's New EU Terrorism Situation and Trend Report 2021.' *Press Release*, 22 June, 2021: https://www.europol.europa.eu/newsroom/news/terrorists-attempted-to-take-advantage-of-pandemic-says-europol%E2%80%99s-new-eu-terrorism-situation-and-trend-report-2021.

[64] Hamish Goodall. 'Home Affairs Minister Karen Andrews Warns Extremists Radicalised Online During Pandemic.' *7News*, 14 September, 2021: https://7news.com.au/sunrise/on-the-show/home-affairs-minister-issues-karen-andrews-warn-extremists-radicalised-online-during-pandemic-c-3956569.

[65] Jessica Elgot, Vikram Dodd and Jamie Grierson. 'More People May Have Self-radicalised Online in Pandemic, Warns Minister.' *The Guardian* (UK), 16 November, 2021: https://www.theguardian.com/uk-news/2021/nov/16/more-people-may-have-self-radicalised-online-in-pandemic-warns-minister.

[66] See, for example, Elias Visontay. 'Far Right "Exploiting" Anger at Lockdowns to Radicalise Wellness Community, Police Say.' *The Guardian*, 24 February, 2021: https://www.theguardian.com/australia-news/2021/feb/25/far-right-exploiting-anger-at-lockdowns-to-radicalise-wellness-community-police-say.

[67] Moonshot Team. 'COVID-19: The Impact of Social Distancing on Engagement with Violent Extremist Content Online in the United States: Initial Observations.' *Moonshot*, 14 April, 2020: https://moonshotteam.com/social-distancing-white-supremacy/.

[68] ISD Global. 'COVID-19 Disinformation Briefing No. 2.' *ISD*, 9 April, 2020: https://www.isdglobal.org/wp-content/uploads/2020/06/COVID-19-Briefing-02-Institute-for-Strategic-Dialogue-9th-April-2020.pdf.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011480

A June 2020 study of Canadian users showed a "statistically significant increase in searches for violent far-right extremist content" across four cities in the six-week period from the imposition of stay-at-home orders versus prior.[69] Davies, Wu, and Frank's analysis of seven online extremist forums found that posting behaviour on both violent right-wing extremist and incel forums increased significantly following the declaration of the pandemic, but that comparable trends were not observable on left-wing or jihadist forums.[70] Similar findings were made by researchers studying German-speaking right-wing, left-wing, and Islamist extremist activity across a range of mainstream and more fringe platforms across 2020.[71]

EU Member States too reported to Europol an increase in online transnational right-wing activities.[72] Researchers made similar findings. Between the beginning of March and the end of September 2020, German-speaking far-right actors—including those based not just in Germany, but also Austria and Switzerland—on Facebook, Twitter, and YouTube grew their audience an average of 18%.[73] This increase was even more prodigious on Telegram where German language right-wing extremist channels grew by almost 350% between the beginning of the pandemic and September 2020. Important to note here is that this growth was not evenly distributed across these far-right actors but dominated by far-right conspiracy theorists.[74]

Likewise, in Ireland the extreme right made increasing use of Telegram to organise, recruit, and spread disinformation in 2020 and into 2021. The administrators of these Telegrams channels used language and tactics commonly seen in the US and international far-right movements. One in ten messages posted by Irish anti-lockdown and COVID-19 conspiracy theory Telegram channels originated from far-right sources. Anti-mask rallies in Dublin were heavily promoted by far-right channels on Telegram. These same channels were used to promote disinformation that an incident which occurred during one such protest was a staged 'false flag' operation to "smear the protesters."[75] Anti-Semitic slurs were also used to describe counter-protestors.

As regards overall right-wing narrative trends, extreme right users, groups, and movements sought to integrate COVID-19 and responses to it into their existing narratives. Reminiscent of previous pandemics, people were searching for explanations, especially who to hold responsible for events, and the online far- and extreme right had readymade answers for them.[76] In 2021, right-wing extremists continued to exploit the pandemic to support

---

[69] Moonshot Team. 'COVID-19: Increase in Far-right Searches in Canada.' *Moonshot*, 8 June, 2020: https://moonshotteam.com/covid-19-increase-in-searches-for-violent-far-right-content-in-canada/

[70] Garth Davies, Edith Wu, and Richard Frank. 2021. 'A Witch's Brew of Grievances: The Potential Effects of COVID-19 on Radicalization to Violent Extremism.' *Studies in Conflict & Terrorism* [Online First].

[71] Jakob Guhl and Lea Gerster. 2020. *Crisis and Loss of Control: German-Language Digital Extremism in the Context of the Covid-19 Pandemic* (London: ISD): https://www.isdglobal.org/isd-publications/crisis-and-loss-of-control-german-language-digital-extremism-in-the-context-of-the-covid-19-pandemic/.

[72] Europol. *European Union Terrorism Situation and Trend Report 2021* (TE-SAT) (The Hague: Europol), p. 27: https://www.europol.europa.eu/activities-services/main-reports/european-union-terrorism-situation-and-trend-report-2021-tesat.

[73] Guhl and Gerster, *Crisis and Loss of Control*, p.13.

[74] *Ibid.*, p.14.

[75] Aine Gallagher and Ciaran O'Connor. 2021. *Layers of Lies: A First Look at Irish Far-Right Activity on Telegram* (London: ISD): https://www.isdglobal.org/isd-publications/layers-of-lies/.

[76] Davies, Wu, and Frank, 'A Witch's Brew of Grievances.'

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

longstanding anti-Semitic conspiracy theories, anti-immigration and anti-Islam rhetoric, and accelerationist narratives.

Both Germany and France saw a significant rise in anti-Semitic content online during the pandemic.[77] Several of the anti-Semitic narratives relate to the COVID-19 pandemic itself, ranging from conspiracy theories presenting vaccines as a Jewish plot to sterilise or control populations to representations of Jewish people as unhygienic or as themselves a 'virus.' Other older anti-Jewish tropes have also proliferated online in 2020 and 2021, such as Jews ruling international financial, political, and media institutions, and Holocaust denial.

A majority of this anti-Semitic—and other already-described content—is non-violent and not obviously illegal under French and German law, which poses a challenge for technology companies and governments in line with other 'legal but harmful' content.[78] On the other hand, as pointed out by Europol in their TE-SAT 2021, which rounds-up extremist and terrorist activity EU-wide in 2020:

> In general, the COVID-19 pandemic and the ensuing economic and social crises have contributed to polarisation in society, causing attitudes to harden and increasing acceptance of intimidation, including calls to commit violent acts. Expressions of social dissatisfaction increased, both online and offline, with social media playing a facilitating and mobilising role, as well as the proliferation of disinformation and conspiracy theories.[79]

As demonstrated by the figures supplied above, this polarisation created opportunities for right-wing extremist users, groups, and movements to reach and mobilise audiences beyond their traditional supporter circles.

Two ways in which this continued to be facilitated in 2021 were via algorithmic amplification of extreme right content and extreme right users' efforts to cloak their content to avoid takedown and, particularly, the overlap of these: the algorithmic amplification of borderline content. Algorithmic amplification refers to "any process that increases the circulation rate of any particular information or content online."[80] The role of major Internet companies recommender algorithms in facilitating extreme right radicalisation has been the subject of research dating back to at least 2014.[81] These concerns received added impetus in 2021, with Facebook whistle-blower Frances Haugen's assertion that the company's algorithms were "dangerous,"[82] along with the emergence of information about an internal 2019 Facebook study where a fake India-based account was created and studied to see what types of content

---

[77] Milo Comerford and Lea Gerster. 2021. *The Rise of Antisemitism Online During the Pandemic: A Study of French and German Content* (London: ISD), p.9: https://www.isdglobal.org/isd-publications/the-rise-of-antisemitism-during-the-pandemic/.

[78] *Ibid*.

[79] Europol, *TE-SAT 2021*, p.28.

[80] Victoria Jordan, Kristin Thue and Jacopo Bellasio. 2021. *Malign Use of Algorithmic Amplification of Terrorist and Violent Extremist Content: Risks and Countermeasures in Place* (Brussels: RAN PS), p.8.

[81] Derek O'Callaghan, Derek Greene, Maura Conway, Joe Carthy, and Pádraig Cunningham. 'Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems.' *Social Science Computer Review* 33(4).

[82] Karen Hao. 'The Facebook Whistleblower Says its Algorithms are Dangerous. Here's Why.' *MIT Technology Review*, 5 October, 2021: https://www.technologyreview.com/2021/10/05/1036519/facebook-whistleblower-frances-haugen-algorithms/.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011482

it was recommended. It was found that within three weeks the fake account's newsfeed was inundated with "hate speech, misinformation and celebrations of violence."[83]

Extreme right users themselves also have a hand in this. They are cognisant that outright hate, extremist, and terrorist content will generally be removed from major platforms, so they increasingly produce and circulate content that very closely skirts, but does not contravene platforms' content moderation policies.[84] Such content, often referred to as 'borderline,' intentionally uses ambiguous language, and takes the form of 'humour,' memes, and similar, and is difficult for automated content moderation tools to identify and for human content moderators to pass judgement on too. Such borderline content is nonetheless increasingly subject to scrutiny and takedown. An example, is the January 2021 removal by Amazon of 92—largely self-published—books containing Holocaust denial that, until their removal, had been 'recommended' (i.e., algorithmically amplified) by Amazon's systems.[85]

EXTREME RIGHT USERS' CURRENTLY PREFERRED ONLINE PLATFORMS, SPACES, AND TOOLS

Fringe and Mainstream Social Media Platforms

It can be useful to distinguish between major, mainstream social media platforms and the extreme right activity thereon and a diversity of more fringe platforms that host increasing amounts of right-wing extremist content, due at least in part to its accelerating takedown by major platforms. Some major platforms too remain core nodes in the online extreme right ecosystem.

Even prior to Christchurch, never mind 6 January, 2021, the increasing inhospitableness of major social media and other online platforms to extreme right content and activity resulted in far-right—largely US—activists establishing their own platforms that welcome, indeed encourage, just such content and activity. Many of these platforms had their user numbers considerably boosted in 2020 and into 2021.

Between the last week of July 2020 and the final week of July 2021, visits to MeWe jumped nearly 300% to almost 1.5 million weekly visits. Gab grabbed an even larger slice of visitor numbers, which increased 400% in the same period to nearly 3 million visits per week. As already mentioned, Parler was the outlier site amongst those favoured by right-wing users, with its visitor numbers halved to c.200,000 weekly in the same period.[86]

Just because these are US platforms does not mean that they are restricted to US users. Most of these platforms have sizeable European user numbers, especially from English-speaking

---

[83] Sheera Frenkel and Davey Alba. 'The Facebook Papers: In India, Facebook Grapples With an Amplified Version of Its Problems' *The New York Times*, 9 November, 2021: https://www.nytimes.com/2021/10/23/technology/facebook-india-misinformation.html.

[84] For example, YouTube identifies borderline content as "content that comes close to, but doesn't quite violate our Community Guidelines." See https://blog.youtube/inside-youtube/on-youtubes-recommendation-system/.

[85] Mathilde Froth. 'Amazon Removes 92 Books Promoting Holocaust Denial.' *The Jewish Chronicle*, 26 January, 2021: https://www.thejc.com/news/uk/amazon-removes-92-books-promoting-holocaust-denial-1.511149.

[86] Harwell, 'Rumble, a YouTube Rival Popular with Conservatives, Will Pay Creators.'

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011483

countries (i.e., Ireland, UK), but also others. For example, the fringe right-wing alternative social media platform Gab continues to host numerous European-focused groups (see Table 1). The listed groups are all fairly active, and discussions occur mostly in their respective languages (i.e., English, French, German, Dutch, Swedish). While GAB France is only for France-related news and French users, *Gab Francais* is targeted at all francophone users. IN 2021, conspiracy theories around COVID-19 (e.g., #plandemic) and agitation against COVID measures (e.g., anti-lockdowns, anti-mask, anti-vaccine) are prominent features of many of these groups, as is a variety of hate content. Outlinking to and cross-posting from other fringe platforms, such as Bitchute, Gettr, and Rumble, but also major platforms, such as Telegram and YouTube, are also apparent.

Table 1. Selection of Europe-focused Gab Groups and their Membership Numbers

Group Name	*Number of Members*	*URL*
#Brexit GB	28.1k members	https://gab.com/groups/730
GAB France	11k members	https://gab.com/groups/2085
DEUTSCH	9k members	https://gab.com/groups/467
Gab Francais	9k members	https://gab.com/groups/2041
NederlandGAB nl	6.9k members	https://gab.com/groups/2896
Swegab	6.6k members	https://gab.com/groups/2523
FREEDOM-UK-USA and Europe.	2.1k members	https://gab.com/groups/4218

As already pointed out, a major platform that continues to host large amounts of right-wing extremist content is Telegram. This includes users, groups, and channels associated with a number of proscribed right-wing terrorist organisations, such as Atomwaffen Division and Proud Boys. Another major platform that gained enormously in popularity, particularly amongst young people, during the pandemic was TikTok. The Chinese-owned video-sharing platform established in 2016 now has an estimated 1 billion users. TikTok came under scrutiny this year for its uneven response to right-wing extremist content in its users, since summer 2021, up-to-three-minute videos. (The previous maximum length of a Tik Tok video was 60 seconds).

Searches, using a lengthy list of right-wing-related keywords, in the period 4 to 30 June, 2021, found 1,030 videos from 491 TikTok accounts featuring "hateful and extremist content."[87] Footage related to the 2019 Christchurch terrorist attack was described as "easily discoverable on TikTok."[88] This took the form of 30 videos that featured support for the attacker, including 13 containing content livestreamed by Tarrant during the course of the attack.[89]

Together three of the top ten most-viewed videos from the same sample of 1,030 videos had 3.5 million views, displayed content originally produced by imprisoned US white supremacist Paul Miller ('Gypsy Crusader').90 Two of the top ten most-viewed videos in the June sample—with 655,800 and 233,000 views respectively—mocked the victims and denied the existence

---

[87] Ciaran O'Connor. 2021. *Hatescape: An In-Depth Analysis of Extremism and Hate Speech on TikTok* (London: ISD), p.10: https://www.isdglobal.org/isd-publications/hatescape-an-in-depth-analysis-of-extremism-and-hate-speech-on-tiktok/.

[88] *Ibid.*, p.5.

[89] *Ibid.*, p.21.

[90] *Ibid.*, p.36

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

of the 1995 Srebrenica genocide and the Holocaust.91 A separate analysis also appeared this year evidencing the active presence of US far-right militia groups on TikTok.92

New Right-wing Oriented Social Media Platform(s)

GETTR was launched by former US President Trump aide Jason Miller on 4 July, 2021 with the aim of providing the answer for US conservatives who say they are battling censorship on more mainstream platforms. In this manner it resembles Parler and Gab before it and like those platforms it's user interface closely resembles that of Twitter. In addition to American far-right figures such as Steve Bannon and Marjorie Taylor Greene, the platform has attracted the attention of several far-right movements, including QAnon and the Proud Boys. In Europe, French far-right figures such as Alain Soral, Daniel Conversano, Stéphane Ravier, and Eric Zammour all have accounts on the platform. Conversano, Ravier, and Zammour are active users, but with widely varying follower counts: Ravier has 69.5k, Zammour 35.2k, but Conversano only c.3,200.

Former US President Donald Trump is not on Gettr; instead, on 20 October, 2021, he announced plans to launch yet another new platform called TRUTH Social. It is mentioned here due to the high likelihood of it attracting a variety of types of right-wing extremist users, including those based in Europe. The truthsocial.com homepage describes itself as America's "Big Tent" social media platform that "encourages open, free, and honest global conversation without discriminating against political ideology." It is unclear whether the site will ever be operational, however. The platform appears to have missed its own November deadline to launch an invitation-only beta version;[93] the share price of the SPAC company, which Trump's social media firm plans to merge with has since plummeted; and Trump's social media firm is now seeking an additional US$1 billion in funding.[94]

Online Gaming

Online gaming is a core component of the online entertainment industry and has spawned an array of platforms, including not just game hosting services but a variety of adjacent spaces that have been developed for use by gamers.[95] A core concern of those focused on extremist and terrorist-related activity on these platforms is that their users skew young and so may be more vulnerable to radicalisation than older users. For example, a 2020 report from the European Union Counter-Terrorism Coordinator argued that there is a "huge target audience for radicalisation on gaming platforms, especially among young people, who tend

---

[91] *Ibid*., p.16 and p.36

[92] Olivia Little. 'TikTok is Prompting Users to Follow Far-right Extremist Accounts.' *MediaMatters*, 26 March, 2021: https://www.mediamatters.org/tiktok/tiktok-prompting-users-follow-far-right-extremist-accounts.

[93] Christina Milkie and Dan Mangan. 'Trump SPAC Social Media Company Appears to Miss its First Product Deadline.' *CNBC*, 1 December, 2021: https://www.cnbc.com/2021/12/01/trump-spac-social-media-company-appears-to-miss-first-deadline.html.

[94] Krystal Hu. 'Exclusive: Trump's Social Media Venture Seeks $1 Billion Raise – Sources.' *Reuters*, 1 December, 2021: https://www.reuters.com/markets/us/exclusive-trumps-social-media-venture-seeks-1-billion-raise-sources-2021-12-01/.

[95] Suraj Lakhani. 2021. *Video Gaming (Violent) Extremism: An Exploration of the Current Landscape, Trends, and Threats* (Brussels: RAN Policy Support), p.3.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011485

to be more vulnerable to radicalisation and recruitment."[96] Europol's 2021 TE-SAT report points to a similar trend, observing "[t]his might help explain the increasingly young suspects arrested for right-wing terrorism and extremism."[97]

Although not new, 2021 saw continued reliance by the extreme right on gaming platforms such as Steam, Discord, Dlive, and Twitch. The Institute for Strategic Dialogue conducted. A key finding of a series of studies on extreme right content and activity across these platforms in 2021 by ISD was that the vast majority of extreme right users and communities observable on these platforms were not connected with a specific extremist group or groups but were instead active in the broader online extreme right movement(s) or 'scene(s).'[98]

Valve-owned Steam was launched in 2003 as a digital game distribution service. Today, it describes itself as the "ultimate destination for playing, discussing, and creating games." In addition to installation and automatic updating of games, it provides server hosting, video streaming, and social networking services, including friends lists and groups, in-game voice and chat functionality, and cloud storage. In research conducted in 2021, Vaux and colleagues found that "the extreme right use Steam as a hub for individual extremists to connect and socialise...[and that] Steam seems to have an entrenched and long lasting extreme right community."[99]

Going to users being active in what might be termed the extreme right online gaming 'scene' broadly, only two Steam communities expressly affiliated themselves with a right-wing extremist group. One community affiliated with the Nordic Resistance Movement (NRM) and contained links to the NRM's official website. This channel was relatively small, with only 87 members but appeared to be active.[100] The other community was named after Misanthropic Division (MD), a Russian extreme right group now active in several countries, including Germany, the UK, and Ukraine. This channel had little public content but more members, a number of which used extreme right terminology in their screen names.[101]

Like Steam, Discord (estbd. 2015) was originally aimed at those wishing to communicate when playing multi-player gamers. Currently, it has real time text, voice, and video chat functionality, with users also able to share a variety of file types. Recent analysis of the type of extreme right content and interactions prevalent on Discord's 'servers' (i.e., chat rooms) indicated these primarily hosted racist trolls who drew on the white nationalist and supremacist forum cultures of sites such as 4chan and 8kun, along with neo-Nazi content. While none of the servers were directly affiliated with violent extremist groups, users did

---

[96] EU Counter-Terrorism Coordinator. 'Online Gaming in the Context of the Fight Against Terrorism.' *Council of the European Union*, 6 July, 2020, p.4: https://data.consilium.europa.eu/doc/document/ST-9066-2020-INIT/en/pdf.

[97] Europol, *TE-SAT 2021*, p.91.

[98] Jacob Davy. 2021. *Gamers Who Hate: An Introduction to ISD's Gaming and Extremism Series* (London: ISD): https://www.isdglobal.org/isd-publications/gamers-who-hate-an-introduction-to-isds-gaming-and-extremism-series/.

[99] Pierre Vaux, Aoife Gallagher, Jacob Davey. *Gaming and Extremism: The Extreme Right on Steam* (London: ISD), p.4: https://www.isdglobal.org/isd-publications/gaming-and-extremism-the-extreme-right-on-steam/.

[100] Vaux *et al.*, *Gaming and Extremism*, p.6.

[101] *Ibid*.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011486

promote the graphics and activity of the proscribed terrorist groups Atomwaffen Division and Sonnenkrieg Division.[102]

The past year has seen a variety of gaming platforms take action against right-wing extremist users and content. Discord was once viewed as a haven for white nationalists. It began to turn this around from 2017 however, following organisers of and attendees at the Unite the Right rally in Charlottesville, USA, making heavy use of the platform. In April 2021, Discord reported that it had banned more than 2,000 extremist communities.[103]

In terms of streaming platforms, the blockchain-based DLive (estbd. 2017) was found to be home to extreme right influencers, including white supremacists and white nationalists, promoting a range of talking points. In response to increased content moderation by the platform in 2021 some of these influencers have instead opted for alternatives such as Trove and Odysee, which have yet to impose content moderation practices.[104]

Twitch is a larger streaming platform, which launched in 2011 and was bought by Amazon in 2014. Twitch, the primary purpose of which is live streaming by gamers of themselves playing games, entered the headlines when the 2019 Halle shooter livestreamed his attack via the platform. In 2021, far- and extreme right Twitch content was found to be focused on the promotion of conspiracy theories, misogyny, and white supremacist worldviews.[105] In general, it doesn't appear that a large right-wing extremist community of content creators currently exists on the platform.

Cryptocurrency

Extremists and terrorists have a lengthy history of online fundraising, including soliciting donations and sale of merchandise via websites and social media.106 While it is not wholly clear how much money is involved, extremists and terrorists have been accepting funds in cryptocurrency for some time too. In 2018, infamous neo-Nazi troll Andrew Auernheimer, a.k.a Weev, encouraged the use of Monero, as it "best maintains our privacy." 107 Extreme right users and groups donated and received cryptocurrency throughout 2020 and into 2021. In December 2020 a terminally ill French donor[108] sent at least a dozen far-right individuals and groups over US$500,000 in Bitcoin. Recipients included Nick Fuentes, the Daily Stormer, and a French Holocaust denier. Fuentes, an alt-right personality and leader of the so-called 'Groyper Army,' was present at the Capitol Hill Riots though he maintains he only protested

---

[102] Davy, *Gamers Who Hate*, p.6.

[103] Bobby Allyn. 'Group-Chat App Discord Says It Banned More than 2,000 Extremist Communities.' *NPR*, 5 April, 2021: https://www.npr.org/2021/04/05/983855753/group-chat-app-discord-says-it-banned-more-than-2-000-extremist-communities.

[104] Elise Thomas. 2021. *Gaming and Extremism: The Extreme Right on DLive* (London: ISD), p.7: https://www.isdglobal.org/isd-publications/gaming-and-extremism-the-extreme-right-on-dlive/.

[105] Ciaran O'Connor. 2021. Gaming and Extremism: The Extreme Right on Twitch (London: ISD): https://www.isdglobal.org/isd-publications/gaming-and-extremism-the-extreme-right-on-twitch/.

[106] Michael Jacobson. 2010. 'Terrorist Financing and the Internet.' *Studies in Conflict & Terrorism* 33(4).

[107] Julia Ebner. 'The Currency of the Far-right: Why Neo-Nazis Love Bitcoin.' *The Guardian*, 24 January, 2018: https://www.theguardian.com/commentisfree/2018/jan/24/bitcoin-currency-far-right-neo-nazis-cryptocurrencies.

[108] Conor Maloney. 'Unmasking the Deceased Programmer Who Donated 28 Bitcoin to Capitol Hill Rioters.' *Crypto Briefing*, 16 January, 2021: https://cryptobriefing.com/unmasking-deceased-programmer-donated-bitcoin-capitol-hill-rioters/.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011487

outside and did not enter the Capitol building.[109] He received the largest share of the donated bitcoin worth US$250,000.

In their TE-SAT 2021 Europol pointed out that "[t]he use of cryptocurrency for terrorism financing is attractive for organisations whose access to the banking system is increasingly restricted."[110] The Nordic Resistance Movement's website's 'Donation' page lists cash, electronic transfer, and an extensive list of crypto currencies the group can accept, including Bitcoin, Ethereum, Litecoin, and Monero. At least one reason that NRM accepts cryptocurrency donations is, as pointed out on the same page, that the Nordic countries have terminated their bank accounts.[111]

Other sources of cryptocurrency include the already mentioned live streaming platform DLive, which allows users to make donations to other users in the form of cryptocurrency. Between 15 April 2020 and early February 2021, 55 far-right individuals and groups were able to earn a total of US$866,700 via the platform. In response to the Capitol Hill events, DLive indefinitely suspended those who used the platform to livestream their own and others participation and also rescinded those users' access to any cryptocurrency tokens given to them by community members.112

Websites

Extremist and terrorist websites never really went away, they were just overlooked for a decade by researchers and others due to a not unwarranted narrowing of focus to social media platforms and, latterly, messaging applications and adjacent online spaces.

While there is less reliance on websites by extremists and terrorists than there once was, websites remain an important component of the contemporary online extremist and terrorist ecosystem(s) and could re-emerge more strongly with accelerated disruption of extremist and terrorist content and accounts by social media platforms and adjacent services unless providers further down 'the tech stack' (e.g., cloud services, content delivery networks, domain registrars)[113] take more concerted action.

Websites played a key role in events at the US Capitol on 6 January 2021. These included Ali Alexander's StopTheSteal.us, Women for America First's TrumpMarch.com, WildProtest.com—a URL paying homage to the Trump tweet displayed in Figure 2—the Oath Keeper's site, and MarchToSaveAmerica.com.[114]

---

[109] 'Alt-Right Groups and Personalities Involved in the January 2021 Capitol Riot Received over $500K in Bitcoin from French Donor One Month Prior.' *Chainalysis*, 14 January, 2021: https://blog.chainalysis.com/reports/capitol-riot-bitcoin-donation-alt-right-domestic-extremism.

[110] Europol, *TE-SAT 2021*, p.33.

[111] See https://nordicresistancemovement.org/donations/.

[112] Peter Stone. 'US Far-right Extremists Making Millions Via Social Media and Cryptocurrency.' *The Guardian*, 10 March, 2021: https://www.theguardian.com/world/2021/mar/10/us-far-right-extremists-millions-social-cryptocurrency.

[113] For more on this, see Maura Conway and Seán Looney. 2021. *Back to the Future? Twenty First Century Extremist and Terrorist Websites* (Brussels: RAN Policy Support), pp.'s 26–30.

[114] DFR Lab, '#StopTheSteal.'

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011488

In the EU and wider European context, the neo-Nazi Nordic Resistance Movement's (NRM) maintains an extensive online presence, including two prominent websites, one in Swedish (www.nordfront.se) and another in English (nordicresistancemovement.org), and a number of local branch sites. While the nordicresistancemovement.org website avoids explicitly violent imagery, it contains text expressing violent ideas. For example, in the item titled 'Simon Lindberg's Thoughts on the Mosque Shootings,' the NRM leader, Lindberg, justified the 2019 Christchurch terrorist attack by reference to falling white birth rates and violent jihadist terrorism, even though Lindberg described the shooting itself as "counter-productive."

Thomas' 2021 research on far-right websites found that rather than being custom-built by professional developers or web designers, the majority were built using open-source technologies. Specifically, 63 of the 100 sites in her dataset were built using WordPress, as was the NRM site.[115] Thomas' figures for the extreme right's use of open-source website development tools and software are out of line with general use of open-source technologies to build websites. Approximately 34% of all websites today use WordPress' content management system; Thomas' figure of 63% hovers around double this.

---

[115] Elise Thomas. 2021. *Open Source, Self Defence: Tackling the Challenge of Extremist Websites and Open Source Tech* (London: Institute for Strategic Dialogue), p.8: https://www.isdglobal.org/wp-content/uploads/2021/08/Open-Source-Self-Defence_v2.pdf.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011489

PART II.
JIHADI ONLINE CONTENT AND ACTIVITY IN 2021

Despite the loss of its 'Caliphate' and a downturn in terrorist attacks in the West, IS was still active globally in both 'real world' and online settings in 2021. Similar to concerns around increased time online leading to some users having increased exposure to extreme right content, more time spent online during the pandemic has probably also increased some users' consumption of and networking around jihadi content.[116]

JIHADI TERRORISM IN 2021

Six violent jihadist terrorist attacks occurred in the West in 2021. On 22 March a 21-year-old man shot 10 people dead at a supermarket in Boulder, Colorado, USA. It is unclear whether the shooter had any links to violent jihadism. Analysis of his Facebook account revealed that while he did mention Islam, there was no indication on his account that suggested he held radical views of any kind.[117] Nonetheless, the shooter was touted as a role model in AQAP's 'Inspire Guide #6: The Colorado Attack' in which he was described as "our hero brother" and a "mujahid," but which claimed no direct link to the shooter or the attack.[118]

On 23 April a female police employee was killed in a knife attack at a police station in the Paris commuter town of Rambouillet. The attacker was shot by police during the attack.[119] The attacker's Facebook posts were exclusively concerned with the defence of the Muslim Community, and the Islamophobia displayed by the French polemicist, and current presidential candidate, Eric Zemmour. Following the October 2020 assassination of Samuel Paty, the attacker had joined a Facebook group titled '*Respectez Mohamed prophète de Dieu*' (Respect Mohamed, Prophet of God), which legitimised violence against those offending Prophet Mohamed.[120]

In May, the pro-al-Qaeda media outlet Thabat published an article in praise of the Rambouillet attacker. The article noted that the stabbing came "just days" after the latest issue of the AQAP magazine Wolves of Manhattan. In that issue an article offered a bitcoin reward "to anyone who can kill a member of the Crusader police in Western countries." On 26 April Thabat had issued a brief post praising the attacker, calling him a "mujahid brother",

---

[116] Europol, *TE-SAT 2021*, pp.'s 26–27.

[117] Coleen Slevin. 'Lawyer: Colorado Shooting Suspect Needs Mental Health Review.' *AP News,* 26 March, 2021: https://apnews.com/article/ahmad-al-aliwi-alissa-1st-court-appearance-boulder-shooting-d870b03a763d8b20fdfa04a26a1379a1.

[118] Mina al-Lami. 'Al-Qaeda Gives Detailed Instructions to "Lone Wolves" in US, West.' *BBC Monitoring,* 1 July, 2021.

[119] 'French Police Station Stabbing: Terror Inquiry into Rambouillet Knife Attack.' *BBC News,* 23 April, 2021: https://www.bbc.co.uk/news/world-europe-56862436.

[120] 'Rambouillet: Ce Que l'On Sait de l'Assaillant Qui a Tué une Fonctionnaire de Police.' *France Info*, 24 April, 2021: https://www.francetvinfo.fr/faits-divers/terrorisme/fonctionnaire-de-police-tuee-a-rambouillet/fonctionnaire-de-police-tuee-a-rambouillet-ce-que-l-on-sait-de-l-assaillant_4384485.html.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011490

without suggesting that he was linked to any particular jihadist group.[121] (Thabat has since been removed from Telegram).

On 3 September six people were stabbed and wounded in a supermarket in Auckland, New Zealand. The attacker was known to authorities as a supporter of IS prior to the attack and was under 24 hour surveillance at the time.[122] He had previously been subject to prosecution under the Terrorism Suppression Act after he had attempted to travel overseas to join IS. Specifically, he had purchased a knife and shown a lot of interest in IS propaganda that encouraged conducting knife attacks in the West.[123]

On 13 October four women and a man were killed and two others wounded when a 37-year-old man used a bow and arrow to attack them in Norway. It was reported that the suspect had converted to Islam and was known to police over fears he had been radicalised.[124] In contrast to the Colorado and Rambouillet attacks there was a limited jihadist reaction online to this attack with no publications or media outlets publishing articles in relation to it. On the pro-IS RocketChat server a few low-profile IS supporters discussed the attack, hoping "the brother" (i.e., attacker) acted on behalf of IS.[125]

On 15 October UK MP for Southend West David Amess was stabbed to death. The attacker had been previously referred to the UK government's Prevent scheme, a programme to stop radicalisation, but he was not considered a subject of interest for security agencies.[126] The incident has been formally declared an act of terrorism,[127] and an early investigation has revealed "a potential motivation linked to Islamist extremism."[128] Similarly to the Norway attack there has been very limited jihadi reaction to this attack. A jihadist news-focused channel, Sanam News Agency, factually reported on the incident, highlighting Amess's membership of the Conservative Party.

Noted here too is the 14 November Liverpool incident, in which an improvised explosive device carried by a taxi passenger blew up, killing him and injuring the driver. The precise nature of this attack is unclear at time of writing. The attacker was an asylum seeker who was reported as being from variously Syria, Jordan, and Iraq, and a convert from Islam to

---

[121] 'Jihadist Outlet Says France Stabbing Followed 'Lone Wolf' Incitement.' *BBC Monitoring,* 4 May, 2021.

[122] '"It Was Hateful, It Was Wrong": Six Injured in NZ Stabbing Attack.' *Al Jazeera,* 3 September, 2021: https://www.aljazeera.com/news/2021/9/3/new-zealand-pm-says-mall-stabbing-a-terrorist-attack.

[123] Tess McClure. 'Auckland Stabbings: Calls for Tighter Terror Laws After Extremist Allowed to Roam Free.' *The Guardian*, 4 September, 2021: https://www.theguardian.com/world/2021/sep/04/auckland-stabbings-calls-for-tighter-terror-laws-after-extremist-allowed-to-roam-free.

[124] 'Kongsberg: Five Dead in Norway Bow and Arrow Attack.' *BBC News,* 14 October, 2021: https://www.bbc.co.uk/news/world-europe-58906165.

[125] 'Limited Jihadist Reaction to Norway Bow and Arrow Attack.' *BBC Monitoring*, 15 October, 2021.

[126] 'Sir David Amess Killing: What We Know About Ali Harbi Ali Who is Being Held on Suspicion of Murder.' *Sky News*, 18 October, 2021: https://news.sky.com/story/sir-david-amess-killing-what-we-know-about-ali-harbi-ali-who-is-being-held-on-suspicion-of-murder-12437757.

[127] UK Counter Terrorism Policing. 'Counter Terrorism Policing Response to the Tragic Murder of Sir David Amess.' *Press Release*, 15 October, 2021: https://www.counterterrorism.police.uk/counter-terrorism-policing-response-to-the-tragic-murder-of-sir-david-amess/.

[128] Peter Walker, Vikram Dodd, Matthew Weaver, and Dan Sabbagh. 'David Amess: MP's Killing Declared a Terrorist Incident.' *The Guardian,* 16 October, 2021: https://www.theguardian.com/uk-news/2021/oct/15/counter-terrorism-police-take-over-inquiry-into-david-amess-killing.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011491

Anglicanism. He had a history of mental health issues, having previously been sectioned under the Mental Health Act due to his behaviour with a knife.[129] Four men were arrested in connection with the explosion but were subsequently released.[130]

Again, this incident received only limited reactions in online jihadi spaces. A handful of IS supporters on Telegram shared a factual news report of the incident, and an obscure pro-IS Rocket.Chat account shared news of the incident with some positive commentary: "All praise to Allah" and a smiley emoji.[131] One factor in the relative lack of online response to the final three attacks is the suspension of jihadi accounts from Telegram in the run-up to the 9/11 anniversary in September.[132] The muddy nature of the information surrounding the Liverpool attack, including the perpetrator's conversion, was definitely also a factor in the muted response to that incident.

Worth noting here, finally, is that internationally, there were many jihadi attacks carried out in 2021. Perhaps the deadliest was the 26 August suicide bomb attack by IS Khorasan Province at Kabul Airport that killed 183 people and injured hundreds more. A large number of fatalities (c.150) also resulted from the overrunning of the town of Palma in Mozambique, by IS's Central Africa Province, between 24 March and 5 April, 2021. Palma was destroyed by the attack, and French energy giant Total SE decided to suspend all operations in the area due to the events.

In terms of online IS content accompanying the attack, they released an initial text statement claiming it. They followed-up with the release of a just-over-one-minute video containing raw footage of the events in Palma on 29 March via its Amaq 'news agency.' The video was distributed via the group's accounts on Telegram. Notable is that while news of the attack on Palma was first reported by international news outlets on 24 March, the IS claim didn't appear until days later. BBC Monitoring suggested that this may have been due to the group facing "communication challenges in getting its propaganda out, possibly due to internet connectivity." (Alternatively, they pointed out, the claim may have been delayed for operational purposes, which has happened previously in respect to Mozambique attacks).[133]

TRENDS IN JIHADI ONLINE CONTENT

IS's loss of territorial control in 2017 negatively affected not only its capacity to orchestrate complex directed attacks in the West but also, as predicted,[134] damaged the group's content production capacities due to the loss of media production facilities, equipment, and personnel. The quantity and quality of official IS online propaganda was therefore in a

---

[129] 'Liverpool Bomb: What We Know So Far.' *BBC News*, 18 November, 2021: https://www.bbc.co.uk/news/uk-59287001.

[130] 'Liverpool Bomb: Four Freed as Suspect Killed in Blast Named.' *BBC News*, 16 November, 2021: https://www.bbc.co.uk/news/uk-england-merseyside-59301708.

[131] 'Limited Jihadist Reaction to Liverpool Blast.' *BBC Monitoring*, 15 November, 2021.

[132] 'Key Pro-al-Qaeda Voices Still Absent Online.' *BBC Monitoring*, 21 October, 2021.

[133] Abdirahim Saeed and Paul Brown. 'Decoding IS's Video of Mozambican Town Assault.' *BBC Monitoring*, 30 March, 2021.

[134] Maura Conway. 2018. *Violent Extremism and Terrorism Online in 2017: The Year in Review* (Dublin: VOX-Pol): pp.'s 1–6: https://www.voxpol.eu/download/vox-pol_publication/YiR-2017_Web-Version.pdf.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011492

continued state of decline in 2021, with text-based communiques being much more prevalent than the release of videos or even still images. It is estimated that an average of one official IS video a month was released in 2021, which is a stark contrast to the estimated 50 videos produced in August 2015. This downturn was somewhat offset by both the sustained production and circulation of supporter content and fallback on 'old' material. Other jihadis, besides IS, particularly al-Qaeda and its affiliates, also continued to produce and circulate online content in the 12-month period under review.

Like extreme right users and groups, jihadi terrorists too sought to use the COVID-19 pandemic for propaganda purposes, with IS interpreting the pandemic as a punishment from God on his enemies, and Al-Qaeda framing the spread of COVID-19 in Islamic-majority countries as "a sign that people had abandoned true Islam, and appealed to Muslims to seek God's mercy by liberating Muslim prisoners, providing for people in need, and supporting jihadist groups."[135] These narratives were much more prominent in 2020 than in 2021 however.

IS's most stable online product remains its Arabic-language al-Naba newsletter. First published in 2014, 50 issues appeared in the period under review here, the first (no. 263) on 3 December, 2020 and the last (no. 313) on 18 November, 2021. In 2021, issues were a standard 12 pages long and each contained two page-length infographics. Both the first and last issue in the period under review had as their frontpage headlines IS activity in Africa, with the December 2020 issue focused on Chad and Nigeria and the November 2021 issue focused on Uganda, Mozambique, and Congo.

As has become commonplace, IS and al-Qaeda, their franchises, and supporters continued to seek to 'inspire' lone actor attacks in the West in 2021. As in previous years too, the aftermath of many of the above-described attacks saw an increase in such calls. Like in 2020, in 2021 IS underlined the 'successes' of its franchises outside of Iraq and Syria, especially in Africa, in an effort to detract from the loss of their 'Caliphate.' The plight of various Muslim populations was also emphasised (e.g., Rohingya), as was that of jihadi prisoners, including IS-affiliated women and their children held in camps in Syria. Al-Qaeda messaging also continued to portray itself as distinct from IS in a variety of ways, including focusing largely—though not exclusively—on 'local' issues.

Fan-produced IS content is not new. IS's October 2018 'Inside the Caliphate #8' video contained clear acknowledgement of the increased amounts of user-generated content (UGC) as distinct from official IS content produced by IS 'fanboys'—and 'fangirls'?—apparent in online settings more than three years ago.[136] The video, which was largely concerned with dos and don'ts for IS's active online *munasirun* or 'supporters,' had the function of both crediting *munasirun* activity while, at the same time, seeking to rein it in somewhat. Times have changed considerably since 2018 however and in 2021 supporter content and activity was viewed by IS as crucial and frequently encouraged.

An important component of this activity was the establishment of new online media outlets by IS supporters; Europol already had this to say about these in 2020:

---

[135] Europol, *TE-SAT 2021*, p.27.

[136] Conway, *Violent Extremism and Terrorism Online in 2018*, pp.'s 7–9.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011493

As official IS media outlets like Nashir News and A'maq News faced increasing suspensions, new media outlets took over the task of delivering IS' daily feed of information. Media outlets supporting IS stepped up their capabilities and attempted to take on the role of official media outlets. Uqab News, for instance, surfaced in 2020 and produced digital media products such as statements and infographics reporting news from the battlefield, thereby mimicking the role traditionally fulfilled by A'maq News.[137]

This activity has continued into 2021, including as a way to avoid takedown from major social media platforms by posing as legitimate news outlets.

Tech Against Terrorism reported seeing "several coordinated efforts by supporter networks of designated terrorist organisations to disseminate content on mainstream platforms under the guise of 'reporting' on current events."[138] These networks' behaviour adapts to the specific platforms on which they are operating aided by their attention to these platforms' Terms of Service and well-developed knowledge of their content moderation decisions more broadly. So, for example, they utilise usernames and profile pictures of a journalistic sort, avoid direct references to terrorist groups, don't display logos associated with the latter, and seek to avoid automated moderation via unusual spellings of words, including the insertion of numbers and other characters.[139]

IS supporters were not the only ones establishing new media outlets in 2021. The al-Fursan (Knights) Media Group and al-Hikma (Wisdom) were both promoted in July by pro-AQAP personality 'Warith al-Qassam' on his Telegram channel, which has since been removed by the company.[140] Another new pro-AQAP media group promoted by al-Qassam was Qurtoba Media. Their debut 14 second video showed what appeared to be a 'martyrdom' operation on a police station in Afghanistan.[141]

In terms of magazines, 2021 saw the release of five new issues of the pro-al-Qaeda women's magazine Ibnat al-Islam (Daughter of Islam) by the Bayt al-Maqdis media outfit. Issue 19 in April, was published predominantly in Arabic but had a section in English and covered topics such as the proposed hijab ban in France, and the plight of Rohingya Muslins in Myanmar. All were very lengthy, routinely running to over 70 pages; display a Gmail email contact address; and were shared via Bayt al-Maqdis's channel on the decentralised Rocket.Chat platform.

2021 also saw the release of the first edition of pro-IS Urdu magazine: *Yalghar* (Invasion). It was attributed to IS's Pakistan Province and claimed to be the first IS magazine in Urdu. Multiple articles concerned IS's rivals, the Afghan Taliban, declaring them stooges of Pakistani intelligence. The longest article, allegedly penned by a female member of IS, relays the difficulties she and her family faced at the time of the group's collapse in Syria in 2019.[142]

---

[137] Europol, *TE-SAT 2021*, p.58.

[138] Tech Against Terrorism. 2021. *Trends in Terrorist and Violent Extremist Use of the Internet | Q1-Q2 2021* (London: Tech Against terrorism), p.6: https://www.techagainstterrorism.org/wp-content/uploads/2021/07/Tech-Against-Terrorism-Q1-Q2-TVEC-Trends-2021.pdf.

[139] *Ibid*.

[140] 'Al-Qaeda Supporter Promotes Two New Jihadist Media Outfits.' *BBC Monitoring*, 31 July, 2021.

[141] 'New Pro-al-Qaeda Media Group Emerges Online.' *BBC Monitoring*, 3 August, 2021.

[142] BBC Monitoring. 'First Edition of Pro-IS Urdu Magazine Released.' *BBC Monitoring*, 4 May, 2021: https://monitoring.bbc.co.uk/product/c202jmcv.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011494

Finally, there was only muted reaction to the video message of al-Qaeda's leader Ayman al-Zawahiri released on 23 November. In it Zawahiri railed against Muslims interacting with the UN, which he described as a betrayal of Islam and abandonment of Sharia. Online jihadists who commented on the video—mostly IS supporters—took it to be veiled criticism of the Taliban, which has been seeking to fill Afghanistan's seat in the organisation. While Zawahiri mentioned the Taliban in the video, it was not in connection with any recent events.[143]

This was only the second video from the al-Qaeda leader in 2021, the first was published on 11 September, 2021 to commemorate the 9/11 attacks, but was focused on condemning Arab governments that have normalised relations with Israel rather than addressing 9/11 itself. The hour-long production was released on 11 September by al-Qaeda's longstanding media outlet al-Sahab via Rocket.Chat. The Arabic-language video was subtitled in English, and transcripts in both languages were provided. A lengthy book by Zawahiri was published online at around the same time and, again, was framed as commemorating 9/11, but it too hardly mentioned the attacks.

JIHADI USERS' CURRENTLY PREFERRED ONLINE PLATFORMS, SPACES, AND TOOLS

As in 2020, IS's online fans struggled to maintain their networks in 2021 in the face of significant disruption by major and a raft of medium platforms. While al-Qaeda and its affiliates and other jihadis were relatively less affected, they too were subject to widespread takedown in 2021. As a result, jihadi networks came to be spread across an even wider array of platforms and online spaces than previously. This dispersal made it increasingly difficult for users to find and follow relevant accounts, channels, and similar and therefore access and communicate about jihadi online content.

As has been ongoing for a number of years, jihadis and their supporters, including especially IS, worked throughout 2021 to regain a sufficient foothold on major mainstream social media platforms, including especially Facebook and Twitter. Often these efforts were coordinated on other platforms and messaging applications and exploited gaps in the targeted companies' moderation processes. So-called 'online raids' also persisted in 2021, with users coordinating to flood the comments' sections of public pages with terrorist content. Not all activity on major platforms was of this nature, however. In December 2020, for example, an imam in France was imprisoned for *apologie du terrorisme* after using TikTok to post extremist videos. In court, the imam described his use of TikTok as a way to seek fame as opposed to a strategy for reaching a young Muslim audience.[144]

Many of probably the same users as coordinating the above-described raids and other activity also made efforts throughout the year to establish themselves on an array of new and/or up-and-coming platforms. Within a short time of GETTR's launch on 1 July, for example, IS supporters had created a raft of new accounts on the platform. Content shared by these accounts included an image of Donald Trump in an IS-style execution and posts

---

[143] For more on online responses to Zawahiri's video, see 'Jihadists See al-Qaeda Message as Criticism of Taliban.' *BBC Monitoring*, 26 November, 2021.

[144] 'How Pakistani Imam Used TikTok in France to Spread Terror.' *France 24 News*, 27 November, 2020: https://www.fr24news.com/a/2020/11/how-pakistani-imam-used-tiktok-in-france-to-spread-terror-world-news.html.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011495

delighting that the platform might present an opportunity similar to IS's Twitter "glory days" when their accounts operated with ease on the platform.[145] This dream appeared short-lived however, with GETTR beginning to suspend pro-IS accounts shortly thereafter.

This is not to say that jihadis aren't still relatively free to act in a variety of other online spaces. Telegram and Rocket.Chat are discussed in more detail below due to their continued importance within the jihadi online ecosystem in the 12-month period under review. Websites too were a feature of the jihadi, especially IS, online scene in 2021. Also, addressed below is cryptocurrency and its role during the year.

Telegram

Official IS and 'fan' online content and activity continues to be most easily accessible via the messaging application Telegram, as do a wide array of other jihadi accounts and content. These users, accounts, and content were nonetheless subject to ongoing disruption by the company throughout 2021.

Table 2. Number of IS Bots and Channels Banned by Telegram Per Month in 2021

Month	*Number of Bots and Channels Banned**
Jan-21	19,672
Feb-21	15,669
Mar-21	15,688
Apr-21	15,254
May-21	18,044
Jun-21	15,069
Jul-21	19,121
Aug-21	20,296
Sep-21	34,286
Oct-21	33,439
Nov-21	25,356
Dec-21**	--------
TOTAL	**231,894**

**** Per data supplied on Telegram's official 'ISIS Watch' Channel***
***** December data not available at time of writing***

In December 2016, Telegram established a dedicated 'ISIS Watch' channel, which provides a running tally of numbers of "ISIS bots and channels banned" by them daily. In November 2019 Telegram, with the aid of Europol, made its first truly concerted effort to delete IS content from their platform. Prior to November 2019, the highest ever number of monthly bans by Telegram of IS bots and channels was 14,531 in December 2018. In November 2019 this tripled to 43,215 and increased again in December 2019 to 56,186. This considerable jump in bans was widely commented upon by IS supporters, researchers, and others and others at the time.

The November 2019 ramp-up in disruption by Telegram of IS activity and content has persisted since, with the new high figure for total number of takedowns in a single month (i.e., 56,858) recorded in April, 2020. See Table 2 for Telegram's 2021 takedown data, which is roughly commensurate with its 2020 activity. The largest number of takedowns in 2021 occurred in September, many of these prior to and immediately after the 9/11 anniversary. Since they began making the information available in 2016, Telegram have reported the takedown of a total just shy of 900,000 IS bots and channels (see Table 3).

[145] Mark Scott and Tina Nguyen. 'Jihadists Flood Pro-Trump Social Network with Propaganda.' *Politico,* 2 August, 2021: https://www.politico.com/news/2021/08/02/trump-gettr-social-media-isis-502078.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011496

Without additional context, it is difficult to tell whether the high numbers of takedowns in 2020 and 2021 are largely due to the company's ramp-up in disruption activity first observed in 2019 and persisted with by them since or increased activity by IS users over the course of the pandemic or some combination of these.

Table 3. Number of IS Bots and Channels Banned by Telegram Per Year

Year	*Number of Bots and Channels Banned**
2016**	2,652
2017	82,789
2018	91,233
2019	176,602
2020	311,703
2021***	231,894
TOTAL	**896,873**

**** Per data supplied on Telegram's official 'ISIS Watch' Channel***
***** December 2016 only***
****** To 30 Nov. 2021 only***

Rocket.Chat

Rocket.Chat has been a node in the jihadi online ecosystem since its establishment in 2015. Widespread disruption has meant that it's now a primary way for jihadist groups and their various media outlets to disseminate propaganda, however. Rocket.Chat is a team chat platform, similar to Slack, which allows users to securely text, share files, video conference, and voice message. Crucially though, and in contrast to other similar platforms, it is open-source and decentralised; this decentralisation makes takedowns very difficult because the platform is not controlled by a single corporation, but by its individual users operating independently (i.e., self-hosting).146

The pro-IS Rocket.Chat server TechHaven launched in late 2018 has, for example, displayed remarkable resilience and stability. The same applies to the pro-al-Qaeda Rocket.Chat server GeoNews, which has been stable since its December 2019 launch. While Rocket.Chat's developers have stated that they have no way of acting against content stored on user-operated servers, in March 2021 the platform announced an internal task force to tip off authorities to any illegal use of its open-source code.147

Websites

Another trend seeking to mitigate against the widespread takedown of jihadi accounts and content by social media and messaging companies and others is an increased reliance by jihadi groups and their supporters, including IS and al-Qaeda, on websites.

2021 has seen the emergence of password-protected 'cloud' websites that enable jihadis to share new and old content via URLs. Two important points made about these platforms by Tech Against Terrorism are that (1) they currently provide jihadists with a relatively stable, centralised location in which to store their content due to the process of taking down cloud

---

146 For a useful explainer on the decentralised web or 'DWeb,' see Zoë Corbyn. 'Decentralisation: The Next Big Step for the World Wide Web.' *The Guardian*, 8 September, 2018: https://www.theguardian.com/technology/2018/sep/08/decentralisation-next-big-step-for-the-world-wide-web-dweb-data-internet-censorship-brewster-kahle.

147 Rocket.Chat. 'Rocket.Chat Announces Internal Task Force to Prevent Future Platform Use by Terrorist Groups.' *Press Release*, 26 March, 2021: https://rocket.chat/press-releases/rocket-chat-announces-internal-task-force-to-prevent-future-platform-use-by-terrorist-groups. See also the discussion of Gab's use of Mastodon and the latter's response in Maura Conway. 2020. *Violent Extremism and Terrorism Online in 2019: The Year in Review* (Dublin: VOX-Pol), p.12: https://www.voxpol.eu/download/vox-pol_publication/Violent-Extremism-and-Terrorism-Online-in-2019-The-Year-in-Review.pdf.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011497

sites being "extremely challenging" and (2) most jihadi cloud sites monitored by them utilise open-source software developed by German company NextCloud.[148]

Furthermore, a number of actors have drawn attention to the increased use of traditional websites by, especially, IS supporters. Many of these websites take an archival form too, hosting vast amounts of official IS content, 'historical' and current, including audio, videos, magazines, and posters, in a wide range of languages, including prominently English and French.

This is not a wholly new issue; as drawn attention to in Europol's *TE-SAT 2021*:

> ...2020 saw the creation of websites and the movement of websites to new domains as a measure to avoid takedowns. Deletion particularly targeted official IS and supporter media outlets. In 2020 a constant turnover of branded media entities supportive of IS was noted, with some disappearing and new ones being introduced, probably in an effort to avoid deletion.[149]

As indicated above, hosting, security, and/or other service providers are plainly refusing service to especially IS supporter websites. This is indicated by the cycling of these sites through multiple and increasingly obscure domains and the relatively short online existence of many. For example, the Elokab IS supporter site has cycled through numerous country code top-level (ccTLD) and more general Internet top-level domains (TLD). Minor changes in spelling, as reflected in the shift in the domain name of the 'elokab' website to 'elokabe,' are also utilised for purposes of avoiding takedown, but still being findable by knowledgeable users.[150] Like the extreme right websites mentioned earlier, many of these sites use the free and open-source version of WordPress for site design and management.

Cryptocurrency

In 2020 cryptocurrency was generally solicited by jihadis not for attack funding purposes, but a variety of other types of activity, including support of camp-based women and children and upkeep of online outlets. While the latter type of solicitations didn't cease in 2021, there was an uptick in requests for funds—and, in at least one case, the payment of funds as, essentially a reward—for more directly attack-related purposes.

2021 saw IS seek to move from Bitcoin to other, more private, cryptocurrencies such as Monero. In April the pro-IS media outfit Afaq released infographics on their official website, which was then shared on Telegram and Rocket.Chat, showing how the sending and receiving of Bitcoin was unsafe. The warning pointed out that Bitcoin logs financial records and transactions on the blockchain and therefore transfers can be tracked. Afaq also cautioned that money transfer sites log IP addresses, purchase of Bitcoin, and cooperate with governments.[151]

In July Afaq began promoting the use of Monero as a safe alternative to Bitcoin due to it being "untraceable." Rather than public instructions on how to buy the cryptocurrency, it

---

[148] Tech Against Terrorism, *Trends in Terrorist and Violent Extremist Use of the Internet*, p.4.

[149] Europol, *TE-SAT 2021*, pp.'s 57–58.

[150] For more on this see, Conway and Looney, *Back to the Future?* pp.'s 28–29.

[151] 'Pro-IS Tech Outfit Warns Supporters About Bitcoin.' *BBC Monitoring*, 19 April, 2021.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011498

said that more detailed explanations on how to do so would be provided privately for people who were endorsed by a pro-IS media group.[152]

Similarly, the second issue of Myanmar based pro-IS group *Katiba al-Mahdi fi Bilad Arakan's* (Battalion of the Saviour in the Land of Arakan) English-language magazine contained a one-page poster, including a QR code, for those wishing to donate using Monero.[153]

An array of jihadists besides just IS and their supporters continued their solicitation of cryptocurrency in 2021.[154] Gaza's al-Qaeda-aligned *Jaysh al-Umma* (Army of the Nation) produced a series of posters listing the type of equipment they needed, how much it cost, and asking for contributions to the cost in Bitcoin.[155]

Cryptocurrency is also now being used as an incentive for action. In December 2020, pro-al-Qaeda media outfit Jaysh al-Malahim al-Electroni (al-Malahim Electronic Army) advertised a course on Bitcoin and other cryptocurrencies on their Rocket.Chat channel. They described it as the "first of its kind in the world of jihadist electronic armies."[156] They also used Bitcoin to incite lone actor killings in the West with a poster in their Wolves of Manhattan magazine.[157] The poster offered one bitcoin, which the magazine valued at US$60,000 to "anyone who can kill a member of the crusader police in Western countries."[158] The same issue also featured a tutorial on cryptocurrency.

DIFFERENTIAL RESPONSES TO TALIBAN ONLINE ACTIVITY

The Taliban took over Afghanistan again in August 2021. When the Taliban were last in power, social media had not yet emerged; today 70% of Afghans have access to a mobile telephone and the Taliban are active across a range of social media and messaging applications though others are less accommodating. Meta-owned Facebook and Instagram, for example, were already taking down Taliban accounts and content prior to the takeover.[159] Twitter, on which the Taliban have been active since 2011 have taken a different view, however.

Twitter are allowing Taliban accounts and content to remain on the platform unless they explicitly contravened the company's Terms of Service, especially as regards glorification of violence. Three prominent Taliban accounts that were still active on Twitter on 1 December 2021 were those of their spokespeople: Zabehulah Mujahid (@Zabehulah_M33; 440.7K followers), Qari Yousef Ahmadi (@QyAhmadi21; 92.3K followers), and Dr. Muhammed

---

[152] 'Pro-IS Tech Outfit Promotes 'Untraceable' Cryptocurrency Monero.' *BBC Monitoring,* 28 July, 2021.

[153] 'Pro-IS Myanmar Group Solicits Cryptocurrency Donations via Magazine.' *BBC Monitoring,* 12 May, 2021.

[154] For more on this topic, see Isaac Kfir. 2020. 'Cryptocurrencies, National Security, Crime and Terrorism.' *Comparative Strategy* 39(2).

[155] 'Gaza Jihadist Group Issues Statement on Jerusalem Events.' *BBC Monitoring,* 12 May, 2021.

[156] 'Pro-al-Qaeda Magazine Incites Killing Police in West.' *BBC Monitoring,* 15 April, 2021.

[157] This may have an unforeseen negative effect for the group as successful terrorist attacks have been associated with lower returns on cryptocurrency; see Pankaj C. Patel and Igor Pereira. 2021. 'The Relationship Between Terrorist Attacks and Cryptocurrency Returns.' *Applied Economics* 53(8).

[158] 'Pro-al-Qaeda Magazine Incites Killing Police in West.'

[159] 'Afghanistan: Facebook Continues Ban of Taliban-related Content.' *BBC News*, 17 August, 2021: https://www.bbc.com/news/business-58239786.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011499

Naeem (@IeaOffice; 277.1K followers). Together these three accounts grew their followers by 70,000+ over the course of October and November 2021.

As regards websites, five websites operated by the Taliban went offline on 20 August, 2021.[160] It remains unclear who or what was actually responsible for the shutdown. While it is widely assumed that San Francisco-based online security provider Cloudflare had a hand in it (i.e., removed their DDoS protections), the company did not respond to journalists' requests for comment nor did they publish an official Blog post on the matter.

Agreed by commentators is that Taliban 2.0 were engaged in the waning months of 2021 in what was, for them, a digital charm offensive on public social media while, at the same time, waging a campaign of intimidation against journalists and others via messaging applications, including Telegram and WhatsApp.[161]

---

[160] For more on these websites, see Conway and Looney, *Back to the Future?*

[161] Amanda Florian. 'The Taliban are Using Private Messaging Apps to Threaten Afghan Journalists.' *Rest of World*, 5 October, 2021: https://restofworld.org/2021/the-taliban-are-using-private-messaging-apps-to-threaten-afghan-journalists/.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

PART III. RESPONSES TO ONLINE VIOLENT EXTREMISM AND TERRORISM IN 2021

Recent years have seen the EU Commission and many national governments—and, in 2021, some US states—around the world developing and implementing legislation to counter online hate, extremist, and terrorist content.[162] The 12-month period under review herein (i.e., 1 December, 2020 to 30 November, 2021) was particularly consequential in this respect. Watershed events were the coming into force of the EU's Terrorism Content Regulation and the release of the so-called 'Facebook Papers'.

Platform regulation received added impetus from Facebook whistleblower Frances Haugen's testimony before a variety of legislative committees in the US and Europe before which she advocated for increased regulation of Internet companies. Despite widespread support for regulation, particularly across the EU, much of the regulatory activity described below has been criticised by scholars and human rights experts, largely on the grounds of impinging on free speech and human rights.[163]

The bulk of this sub-section is focused on legislative responses to online extremism and terrorism in 2021, including attention to not just the EU Regulation, but also bills and acts progressing through the legislative process in the UK, Poland, Australia, New Zealand, and the US states of Florida and Texas. Tech company-focused legislation is not the only response mechanism available to state's for responding to online extremism and terrorism, however. The role of proscription of groups as terrorist and the role of law enforcement, in this case Europol's Internet Referral Unit, in content takedown are also discussed below. Before discussing these instruments though, the most consequential corporate response to online extremism and terrorism in 2021 was the initial round of decisions by Facebook's new Oversight Board, which are described and discussed directly below.

FACEBOOK OVERSIGHT BOARD CASES AND DECISIONS 2021

Facebook's CEO, Mark Zuckerberg, first publicly floated the idea of an Oversight Board in November 2018,[164] with the Board's 20 founding members announced in May 2020, and the Board officially beginning its work in October, 2020. Three of the Board's members hail from EU member states: former Danish Prime Minister, Helle Thorning-Schmidt; the Executive Director of Internet Sans Frontières, Julie Owono (Cameroon and France); and Central European University's Prof. András Sajó (Hungary).[165] The Facebook Oversight

---

[162] For more on this, see Nery Ramati. 2020. *The Legal Response of Western Democracies to Online Terrorism and Extremism* (Dublin: VOX-Pol): https://www.voxpol.eu/download/vox-pol_publication/The-Legal-Response-of-Western-Democracies-to-Online-Terrorism-and-Extremism.pdf.

[163] Tech Against Terrorism. 2021. *The Online Regulation Series: The Handbook* (London: Tech Against Terrorism): https://www.techagainstterrorism.org/wp-content/uploads/2021/07/Tech-Against-Terrorism-%E2%80%93-The-Online-Regulation-Series-%E2%80%93-The-Handbook-2021.pdf.

[164] Kate Klonick. 'Inside the Making of Facebook's Supreme Court.' *The New Yorker*, 12 February, 2021: https://www.newyorker.com/tech/annals-of-technology/inside-the-making-of-facebooks-supreme-court.

[165] For a full list of the FOB members and their backgrounds, see https://www.oversightboard.com/meet-the-board/.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011501

Board (FOB) issued its first five decisions on 28 January, 2021, with four out of the five overturning Facebook's actions regarding the issues submitted.

Seventeen decisions of the FOB were delivered in total during 2021. Fourteen of the cases fell into the hate, extremism, and terrorism realms construed broadly (see Appendix 1). Not all of these cases, their outcomes, and Facebook's responses can be discussed here, but details of all can be found on the FOB website at oversightboard.com. In summary however:

- A significant majority of all the cases decided in 2021, 82%, had hate, extremism, and/or terrorism aspects;
- Hate speech dominated, accounting for six cases, five cases fell into the category of 'Dangerous individuals and organizations,' and three 'Violence and incitement';
- Of the 13 cases in Appendix 1 on which a FOB decision was reached, eight overturned Facebook's original decisions and five of these were upheld;
- Just two of the cases decided in 2021 involved proscribed terrorist organisations. The FOB overturned Fakebook's decision regarding takedown of a discussion of the PKK leader Abdullah Ocalan's imprisonment and Facebook reversed its decision and allowed Hamas-related content to stay up once it was alerted the case was going before the FOB. The FOB agreed with the latter reversal;
- Four of the cases had some European aspect, three related to France and one to the Netherlands;
- Two cases also had misinformation aspects, one in conjunction with violence and incitement in France and the other in conjunction with dangerous individuals and organisations.

The FOB decision attracting the most attention to date was 2021-001-FB-FBR, the decision on the banning of former US President Trump's Facebook account in the wake of 6 January, 2021. The FOB decided to uphold Facebook's decision to lock Trump's account. The Board found that two of Trump's posts regarding the events on Capitol Hill violated Facebook's policies prohibiting praising or supporting violence and that, further, Trump's false claims of voter fraud and calls to action created a risk of violence, particularly given his influence and reach. However, the statement that was released by the Board condemned Facebook's decision to impose an "indeterminate and standardless penalty of indefinite suspension," which was out-of-step with "Facebook's normal penalties": removing the violating content, imposing a time-bound suspension, or permanently disabling the page and account.[166]

The Board stated that going forward, Facebook must justify a proportionate and consistent response that is in line with its rules. It said that "it is not permissible for Facebook to keep a user off the platform for an undefined period, with no criteria for when or whether the account will be restored." The Board also criticised Facebook's decision to refer this case to them as avoiding its responsibilities and therefore "insists that Facebook apply and justify a defined penalty."[167] The Board gave Facebook six months to reexamine the arbitrary decision it made on 7 January. It said that the new decision must consider the severity of the violation, the prospect of future harm, and be proportionate and consistent with

---

[166] Facebook Oversight Board. 2021. 'Case Decision 2021-001-FB-FBR.' *Oversight Board*, p.1: https://www.oversightboard.com/sr/decision/2021/001/pdf-english.

[167] *Ibid.*, p.33.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011502

existing rules. The Board also said that "considerations of newsworthiness should not take priority when urgent action is needed to prevent significant harm."[168]

In the event, Facebook announced on 4 June, 2021 that it was introducing "new enforcement protocols to be applied in exceptional cases such as this" and on the basis of these "suspending his accounts for two years, effective from the date of the initial suspension on January 7 this year." They followed-up by saying:

> At the end of this period, we will look to experts to assess whether the risk to public safety has receded. We will evaluate external factors, including instances of violence, restrictions on peaceful assembly and other markers of civil unrest. If we determine that there is still a serious risk to public safety, we will extend the restriction for a set period of time and continue to re-evaluate until that risk has receded.[169]

Given both the cases it has accepted to-date and the possibility of other major companies following suit, the FOB is plainly an experiment to watch in the online violent extremism and terrorism realms.

PROSCRIPTION

Today a host of Western democratic governments, including Canada, the UK, the US, and others, as well as the EU and UN, produce lists of proscribed terrorist organisations. Outside of the US, most have both a domestic and international focus. These render decisions around what does and does not constitute terrorism content and should therefore be removed from their platforms easier for technology companies.

Some caveats apply, however. First, terrorism designation lists have only recently begun to include right-wing terrorist groups. The first such proscription was of National Action by the UK government in December 2016 and a number of similar designations have been made since (see Table 4). Notably, both Canada and the UK proscribed the US neo-Nazi group Atomwaffen Division, also known as Nationalist Socialist Order as a terrorist group in 2021, due to it being linked with at least five murders, explosions, and hate crimes in the US. The group does not appear to have a prominent physical presence in the UK, but it does have an online presence, and the decision to proscribe promotes international solidarity. It also allows the UK government to freeze and seize assets, disrupt activities, and use immigration powers.[170] Canada also proscribed the Proud Boys, which originated there, in February; they are the only country to have done so. None of the 21 organisations presently appearing on the EU's terrorism designation list is extreme right in its orientation.[171]

---

[168] *Ibid.*, p.5.

[169] Nick Clegg. 'In Response to Oversight Board, Trump Suspended for Two Years; Will Only Be Reinstated if Conditions Permit.' *Facebook Newsroom*, 4 June, 2021: https://about.fb.com/news/2021/06/facebook-response-to-oversight-board-recommendations-trump/.

[170] Hayley Evans. 'All You Need to Know About the UK Proscribing the Neo-Nazi Group Atomwaffen Division.' *Lawfare*, 17 May, 2021: https://www.lawfareblog.com/all-you-need-know-about-uk-proscribing-neo-nazi-group-atomwaffen-division.

[171] Worth noting here, however, is that in May 2016 the European Commission launched the 'Code of Conduct on Countering Illegal Hate Speech Online,' which by 2021 had eight signatory companies: Facebook, YouTube, Twitter, Microsoft, Instagram, Dailymotion, Snapchat, and Jeuxvideo.com. For more on this, go to https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_1135. More recently, the European Internet Forum to Counter Terrorism has developed a non-legally binding list of violent right-wing extremist

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011503

Second, some proscription lists are highly political, including those of some democratic countries, and thus some designations have met with resistance within the international community. Israel's October 2021 designation of six Palestinian human rights and civil society groups as terrorist has been widely criticised, for example.[172] Much less contentious was the US's 2021 designation of, amongst others, IS-DRC and IS-Mozambique as terrorist entities (and their removal of Colombia's FARC from the listing).

Table 4. Designation of Extreme Right Terrorist Groups by Democratic Countries and Supranational Organisations

	UN	*EU*	*US*	*UK*	*Canada*	*Australia*
Blood and Honour	---	---	---	---	✓	---
National Action	---	---	---	✓	---	---
Combat 18	---	---	---	---	✓	---
Sonnenkrieg Division	---	---	---	✓	---	✓
Scottish Dawn	---	---	---	✓	---	---
National Socialist Anti-Capitalist Action	---	---	---	✓	---	---
System Resistance Network	---	---	---	✓	---	---
Feuerkrieg Division	---	---	---	✓	---	---
Atomwaffen Division	---	---	---	✓	✓	---
National Socialist Order	---	---	---	✓	✓	---
Russian Imperialist Movement	---	---	✓	---	✓	---
The Base	---	---	---	✓	✓	---
Proud Boys	---	---	---	---	✓	---

✓ = Designated terrorist group ✓ = Designated under synonym or umbrella group or by affiliation

Source: Adapted from Terrorist Content Analytics Platform's (TCAP) 'Group Inclusion Policy' at https://www.terrorismanalytics.org/group-inclusion-policy.

LAW ENFORCEMENT REFERRAL REQUESTS

The EU's Internet Referral Unit (EU IRU) and a number of countries refer online terrorist content to Internet companies hosting it, that then consider whether the referred content breaches companies ToS or other guidelines and should thus be removed from their platform. Referrals are a voluntary tool and help improve public-private cooperation around online terrorism and responses to it.

Europol organised the first Referral Action Day (RAD) targeting right-wing terrorist online content on 27 May, 2021. Coordinated by the EU IRU, it involved law enforcement from 28 countries, including not just EU member states, but also Australia, Georgia, the UK, and the New York City Police Department. The content referred included violent right-wing extremist and terrorist content; proscribed right-wing groups' official or supportive content; content

---

groups, symbols, and manifestos with the aim of providing guidance to EUIF Internet companies in their content moderation efforts.

[172] Tania Krämer. 'Critics Seek Proof After Israel Designates Palestinian Rights Groups as Terrorists.' *Deutsche Welle*, 26 October, 2021: https://www.dw.com/en/critics-seek-proof-after-israel-designates-palestinian-rights-groups-as-terrorists/a-59623937.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011504

concerning right-wing terrorist attacks, including video of attacks, attackers' online manifestos, and similar.[173]

During the Referral Action Day, the parties involved reviewed the referral process, from detecting terrorist content to its flagging to online service providers (OSPs). A total of 1,038 items were assessed for referral to the relevant Internet companies on the day. In a post-RAD press release, Europol stated that "[t]his first Referral Action Day against right-wing terrorist and extremist online propaganda aimed to discover the sources of Internet activities and develop approaches to combat them. Similar joint actions will continue to be organised in the future."[174]

The EU IRU also organised an online jihadist propaganda RAD on 14 July, targeting propaganda hosted on the US-based Internet Archive, which has been a core node in the online jihadi ecosystem for many years. The day involved law enforcement from Denmark, Germany, Hungary, Portugal, Slovenia, Spain, and the UK finding relevant content and then referring it to the Internet Archive for checking against their ToS. The referred content included a variety of types of material, such as audio, video, and digital magazines, produced by IS, al-Qaeda, and their affiliates.[175]

LEGISLATION

EU Regulation on Addressing the Dissemination of Terrorist Content Online

The Regulation addressing the dissemination of Terrorist Content Online (TCO Regulation) will enable competent authorities of EU Member States to send legally binding orders to hosting service providers to remove or block access to terrorist content. The Regulation will become fully applicable on 7 June 2022.

At the core of the regulation is a requirement that tech companies remove content or disable access to it in all EU member states within one hour of having received a removal order from a competent authority.[176] "Member states will adopt rules on penalties, the degree of which will take into account the nature of the breach and the size of company responsible."[177] This regulation does not apply to content that is journalistic, intended to be educational, artistic, or is for research purposes. It also does not impose a general obligation on companies to monitor content; however, once aware of terrorist content, companies must take measures to address its dissemination. Finally, platforms and competent authorities must publish annual transparency reports that include information on the removal orders and actions taken to address the dissemination of terrorist content.[178]

---

[173] Europol. '1st Referral Action Day Against Right-wing Terrorist Online Propaganda.' *Press Release*, 28 May, 2021: https://www.europol.europa.eu/newsroom/news/1st-referral-action-day-against-right-wing-terrorist-online-propaganda.

[174] *Ibid.*

[175] Europol. 'Jihadist Content Targeted on Internet Archive Platform.' *Press Release*, 16 July, 2021: https://www.europol.europa.eu/newsroom/news/jihadist-content-targeted-internet-archive-platform.

[176] *Ibid.*

[177] *Ibid.*

[178] *Ibid.*

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011505

The regulation has been highly criticised, particularly the one-hour timeframe for removal, which it is feared will infringe on freedom of speech. Some have argued that one hour is insufficient time to undertake such requests and could lead to platforms over-blocking to reduce the risk of receiving a penalty. Other platforms may not have sufficient staffing or resources to meet the timeframe, placing unfair burdens on smaller platforms.[179] Although there is not a legal obligation to use automated tools, use of these tools is likely one of the only methods that will allow platforms to meet the strict timeframe. This raises concerns surrounding the errors and bias that come with using these tools, particularly where there is very little oversight and transparency.[180]

UK Online Safety Bill

In 2019, the UK Government said that it would like the UK to be the safest place in the world to use the internet. This resulted in the Online Safety Bill, of which a draft was published in May 2021.[181] Under the draft Bill, "social media sites, websites, apps and other services hosting user-generated content or allowing people to talk to others online must remove and limit the spread of illegal and harmful content such as child sexual abuse, terrorist material and suicide content."[182]

All the companies that fall under the scope of the proposed legislation would be required to accede to a new duty of care "that what is unacceptable offline will also be unacceptable online." The Bill splits tech companies into two categories. 'Category 1 services' are the largest and most popular social media platforms. These platforms will need to take action against content that is illegal but also content that is lawful but still "harmful." They will have to explicitly state in their terms and conditions how they will address these harms. These companies will also have to publish assessments of their impact on freedom of expression and the actions that they took to mitigate any adverse effects. 'Category 2 services,' so smaller companies, must only take action against illegal content.[183]

Under this legislation, Ofcom, which is the UK's already-established independent broadcasting regulator, will have the power to fine companies up to £18 million or ten percent of their annual global turnover, or block access to the platform, if a platform fails to comply with the proposed duty of care. There is also a deferred enforcement option of

---

[179] Balazs Denes. 'Open Letter on Behalf of Civil Society Groups Regarding the Proposal for a Regulation on Terrorist Content Online.' 9 November, 2020: https://edri.org/wp-content/uploads/2020/11/TERREG_Openletter_Liberties.pdf; United Nations Office of the High Commissioner for Human Rights. 'UN Human Rights Experts Concerned about EU's Online Counter-terrorism Proposal.' Press Release, 12 December, 2018: https://www.ohchr.org/en/NewsEvents/Pages/DisplayNews.aspx?NewsID=24013&LangID=E; Tech Against Terrorism. 'Tech Against Terrorism Response to the EU's Terrorist Content Online Regulation.' *VOX-Pol Blog*, 30 June, 2021: https://www.voxpol.eu/tech-against-terrorism-response-to-the-eus-terrorist-content-online-regulation/.

[180] Fanny Hidvegi. 'Open Letter to EU Parliament on the Terrorism Database.' *Access Now*, 7 February, 2019: https://www.accessnow.org/open-letter-to-eu-parliament-on-the-terrorism-database/.

[181] Department for Digital, Culture, Media and Sport. *Draft [UK] Online Safety Bill*. May, 2021: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/985033/Draft_Online_Safety_Bill_Bookmarked.pdf.

[182] Department for Digital, Culture, Media and Sport. 'Landmark Laws to Keep Children Safe, Stop Racial Hate and Protect Democracy Online Published.' *Press Release*, 12 May, 2021: https://www.gov.uk/government/news/landmark-laws-to-keep-children-safe-stop-racial-hate-and-protect-democracy-online-published.

[183] *Ibid*.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011506

criminal liability on senior managers, which will be able to be implemented at a later date if required.[184]

Previous iterations of the proposal were criticised, first, for a lack of clarity surrounding the definition of "harmful" content and then, in December 2020, updating the definition to include content that has a "psychological impact on individuals," which is difficult to determine. Scholars argue that such unclear definitions will result in platforms having to determine for themselves what content is harmful and could result in over-blocking out of fear of penalties.[185] It could cause confusion for users too.[186]

Other criticisms include the use of Ofcom instead of a new independent regulator due to the new tasks being very different to the existing broadcasting regulatory duties that Ofcom undertakes and therefore requiring different expertise. Some argue, in addition, that a new body would have created a positive symbolic effect to showcase the UK Government's commitment to countering this content; also, that it would have minimised confusion among the public as to who to contact regarding any issues.[187] Finally, there is concern surrounding the severity of the penalties albeit it is claimed that blocking access to sites and senior management liability will only be used when other penalties have been exhausted.

Polish Law on Protecting the Freedoms of Social Media Users

At the beginning of 2021, the Polish Ministry of Justice announced the development of a new law to protect the freedoms of social media users. The law aims to prevent the arbitrary power of these platforms by making it illegal for platforms to delete content that does not break Polish law.[188] Under the proposed new law, users could file a petition to a new electronic court—being referred to as the 'Free Speech Council'—to overturn the decision of the platform and restore the removed content if it did not break Polish law.[189] The 'Free Speech Council' would be made-up of law and new media experts and would have seven days to decide on cases before it.[190] If a social media platform does not restore the content within 24-hours then it could receive a fine of up to €11 million.

There are two main concerns with this draft legislation. The first is that the severity of the fine may incentivise platforms to leave potentially violating content online so that they do

---

[184] *Ibid*.

[185] Sally Broughton Micova and Sabine Jacques. 2019. 'HM Government's Online Harms White Paper.' Submission to public consultation from the Centre for Competition Policy, University of East Anglia: https://research-portal.uea.ac.uk/en/publications/hm-governments-online-harms-white-paper.

[186] Stefan Theil. 2019. 'The Online Harms White Paper: Comparing the UK and German Approaches to Regulation.' *Journal of Media Law* 11(1).

[187] Patrick Bishop, Seàn Looney, Stuart Macdonald, Elizabeth Pearson, and Joe Whittaker. 2019. 'Response to the Online Harms White Paper.' Submission to public consultation from CYTREC, Swansea University: https://www.swansea.ac.uk/media/Response-to-the-Online-Harms-White-Paper.pdf.

[188] Panoptykon Foundation. 2021. 'Polish Law on "Protecting the Freedoms of Social Media Users" Will Do Exactly the Opposite.' *EDRi*, 10 February, 2021: https://edri.org/our-work/polish-law-on-protecting-the-freedoms-of-social-media-users-will-do-exactly-the-opposite/.

[189] *Ibid*; Panoptykon Foundation, 'Polish Law on "Protecting the Freedoms of Social Media Users."'

[190] Magdalena Gad-Nowak and Marcin S. Wnukowski. 2021. 'Polish Freedom of Speech on Social Media Platforms.' *Nat Law Review* 11(340): https://www.natlawreview.com/article/polish-government-to-pass-law-will-allow-it-more-control-over-internet-content-and.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011507

not have to go through this process and face penalties.[191] The second is that the Council will be subject to political influence.[192] With regard to the latter, it is noticeable that the proposed Polish law and the discourse around it are very similar to the proposed Florida and Texas laws introduced by their Republican governors and discussed below.

Australian Online Safety Bill

The Australian Online Safety Bill was introduced in Parliament in February 2021. It takes a similar view to the UK Online Safety Bill that what is unlawful offline should also be unlawful online. Amongst other things, the draft bill:

> [E]stablishes an online content scheme for the removal of certain material; creates a complaints-based removal notice scheme for cyber-abuse being perpetrated against an Australian adult; broadens the cyber-bullying scheme to capture harm occurring on services other than social media; reduces the timeframe for service providers to respond to a removal notice from the eSafety Commissioner; brings providers of app distribution services and internet search engine services into the remit of the new online content scheme; and establishes a power for the eSafety Commissioner to request or require internet service providers to disable access to material depicting, promoting, inciting or instructing in abhorrent violent conduct for time-limited periods in crisis situations.[193]

The Bill has a takedown timeframe of 24 hours, proposes a self-regulatory code of conduct that tech companies must adhere to regarding countering online disinformation, and grants the eSafety Commissioner the power to immediately block websites hosting abhorrent violent or terrorist material, such as the Christchurch attack.[194]

New Zealand Livestreaming Bill and Content Regulatory System Review

Also in February 2021, New Zealand passed a Bill that seeks to update the Films, Videos, and Publications Classification Act 1993 in an attempt to address livestreaming after the Christchurch attack. "The Bill provides additional regulatory tools to manage harms caused by content that is livestreamed or hosted by online content hosts."[195] Under this Bill, livestreaming objectionable content is a criminal offence. It only applies to the individual or group responsible for livestreaming the content however, not the platform that provided the livestreaming service.[196] Under the Bill, the New Zealand Department of Internal Affairs (DIA) would appoint inspectors who could issue takedown notices on objectionable content, which could be appealed if believed to be erroneous. If a person continues to disseminate the content after the issue of a notice then they would be committing an offence under this law.

---

[191] *Ibid*.

[192] Panoptykon Foundation, 'Polish Law on "Protecting the Freedoms of Social Media Users."'

[193] Parliament of Australia. *Online Safety Bill 2021*: https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/Bills_Search_Results/Result?bId=r6680.

[194] Madeleine Gandhi. 'Update on Online Harms: Legislative Reforms in Australia and the UK.' *MediaWrites*, 13 April, 2021: https://mediawrites.law/update-on-online-harms-legislative-reforms-in-australia-and-the-uk/#page=1.

[195] New Zealand Parliament. *Films, Video and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Bill*: https://www.legislation.govt.nz/bill/government/2020/0268/latest/d10731009e2.html.

[196] *Ibid*.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011508

Again, concerns have been expressed over a chilling effect on free speech, also that the proposed legislation could be used against legitimate activism.[197] Some live portrayals of violence may be in the public interest. For example, would footage such as the George Floyd murder have been removed under this law? The major tech platforms responded that the Bill must provide greater clarity moving forward if it is going to be implemented.[198]

In addition to the above, a comprehensive review of content regulation in New Zealand was announced by the Minister of Internal Affairs, Jan Tinetti, on 10 June 2021. The review is being managed by the Department of Internal Affairs, with support from the Ministry for Culture and Heritage. It aims to design and create a new modern, flexible and coherent regulatory framework to mitigate the harmful impacts of content, regardless of how it is delivered, while still protecting and enhancing democratic freedoms, including freedom of expression and freedom of the press.[199]

US States

Not only did federal-level US legislators continue to grow their interest in amending Section 230 of the Communications Decency Act (CDA) in 2021, but two states also introduced their own legislative instruments. Both of the latter go to US conservatives' longstanding belief, in the face of evidence to the contrary, that technology companies are censoring them based on their politics.

In May 2021, Florida's Governor Ron DeSantis signed the Stop Social Media Censorship Act. Only months after former US President Trump was suspended from a number of platforms, the Act seeks to stop tech platforms from suspending the accounts of political candidates prior to elections and would also make it easier for users and the Florida Attorney General to legally challenge tech companies' moderation decisions. It states that platforms will be fined up to $250,000 per day if found to violate the Act. In addition to being criticised as unconstitutional, the Act is also criticised as conflicting with the CDA's Section 230, which provides platforms with immunity from liability over their content moderation decisions.[200]

In September 2021, Texas introduced a bill with broadly similar purposes. The proposed Texas law would apply to social media companies with more than 50 million monthly active users in the United States, including Facebook, Twitter, and YouTube. It would require them to produce reports on illegal or potentially policy-violating content, as well as build a complaint system via which users could challenge companies' content moderation decisions or flag illegal activity. In addition, users and the Texas Attorney General could file suit if they believed that tech companies wrongfully banned them. "It is now law that

---

[197] Cameron Sumpter. 'Collaboration and Legislation: Confronting Online Violent Extremism from New Zealand.' *GNET*, 10 May, 2021: https://gnet-research.org/2021/05/10/collaboration-and-legislation-confronting-online-violent-extremism-from-new-zealand/.

[198] *Ibid.*

[199] For more on this, go to the official New Zealand government page on the Review at https://www.dia.govt.nz/media-and-online-content-regulation.

[200] Gilad Edleman. 'Florida's New Social Media Law Will Be Laughed Out of Court.' *Wired*, 24 May, 2021: https://www.wired.com/story/florida-new-social-media-law-laughed-out-of-court/.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011509

conservative viewpoints in Texas cannot be banned on social media," Texas Governor Greg Abbott prematurely stated.[201]

In June, a US federal court judge blocked the Florida law from taking effect, suggesting it would be found unconstitutional after tech industry groups brought a challenge. The Florida Governor's administration appealed the judge's ruling. Tech industry groups have criticised the Texas law too, and the industry is likely to bring a similar challenge in Texas as they did in Florida to prevent it from taking effect.

---

[201] Cat Zakrzewski. 'Texas Governor Signs Bill Prohibiting Social Media Giants From Blocking Users Based on Viewpoint.' *Washington Post*, 9 September, 2021: https://www.washingtonpost.com/technology/2021/09/09/govgregabbott-social-media-censorship-bill/.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011510

CONCLUSION

Overall, the number of right-wing and jihadi terrorist attacks in the West were down during the pandemic, partly due to restrictive measures limiting the chance for offline gatherings. However, there was continued high availability of extreme right content online during the year, despite a concerted crackdown associated with the 6 January, 2021 events on the US's Capitol Hill. Official IS online content production and distribution continued to suffer and a concomitant increase in 'fan' content and activity was observable.

In terms of the narratives displayed in this content, the Covid-19 pandemic and associated measures remained a core component of content across the extreme right spectrum; in contrast, and unlike at the beginning of the pandemic, jihadi online content was largely focused elsewhere in 2021, including an IS concentration on Africa and continued jihadi encouragement of lone actor attacks in the West.

Sustained disruption of extreme right and jihadi content across not just major but a raft of medium-sized platforms—and a variety of other online services—meant that shifts to ever more niche platforms were discernible across both communities. The pace of government regulation efforts also stepped-up in 2021, with a raft of measures either introduced or continuing their progress through the legislative process during the year.

Finally, 12 things to watch for in the online and extremism and terrorism realms in 2022 are outlined in Box 4.

Box 4. Online Extremism and Terrorism: What to Watch for in 2022

- Continued extreme right exploitation of COVID-19, including online to offline trajectories;
- Increasing online and 'real world' ideological and aesthetic crossover, both between ideological 'sets' (e.g., incels-extreme right, QAnon-extreme right, ecology-extreme right) and across ideologies (e.g., far-right-Salafis);[202]
- Continued targeting of women (e.g., QAnon via beauty and wellness online influencers and communities) and youth (e.g., via gaming and adjacent platforms);
- More reliance on audio (e.g., Clubhouse, Spotify, Twitter) by extremists and terrorists due to their insufficient content moderation guidelines, personnel, and tools;
- Similarly to websites, an increasing reliance on resurfaced 'old' technology, such as email newsletter services (e.g., Substack), by extremists and terrorists due, again, to these services relative inattention to content moderation;
- The emergence of new online platforms and services attractive to extremists and terrorists but with no strategies in place to deal with these—and other—harms;
- Increasing exploitation of the federated or decentralised web (i.e., the 'Dweb');
- Continued financial profiting by, especially, extreme right influencers from their online activity;
- Increasing reliance on, especially privacy-preserving, cryptocurrency and other online payment mechanisms for fundraising, donations, and payments;
- The implementation of relevant new legislation, especially the EU's TCO Regulation;
- The progress of relevant measures already-in-train through the legislative process, especially the EU's Digital Services Act (DSA);
- The tabling of relevant new regulatory measures globally, including outside of the Western democratic sphere (e.g., India).

[202] See, for example, Moustafa Ayad. 2021. *Islamogram: Salafism and Alt-Right Online Subcultures* (London: ISD): https://www.isdglobal.org/wp-content/uploads/2021/11/Islamogram.pdf.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011511

APPENDIX 1

Overview of 14 Facebook Oversight Board (FOB) Cases and Decisions with Hate, Extremism, and/or Terrorism Components 2021			
FOB-supplied Alpha-Numerical Identifier	*2021 Decision Date*	*FOB-assigned Tags*	*FOB Decision*
1. 2020-002-FB-UA	28 Jan.	Myanmar, France, China, **Hate speech**, Politics, Religion, Violence	"**Overturned** Facebook's decision to remove a post under its Hate Speech Community Standard. The Board found that, while the post might be considered offensive, it did not reach the level of hate speech."
2. 2020-003-FB-UA	28 Jan.	Armenia, Azerbaijan, **Hate speech**, Culture, Discrimination, Religion	"**Upheld** Facebook's decision to remove a post containing a demeaning slur which violated Facebook's Community Standard on Hate Speech."
3. 2020-005-FB-UA	28 Jan.	United States, **Dangerous individuals and organizations**, Politics	"**Overturned** Facebook's decision overturned Facebook's decision to remove a post which the company claims violated its Community Standard on Dangerous Individuals and Organizations. The Board found that these rules were not made sufficiently clear to users."
4. 2020-006-FB-FBR	28 Jan.	France, **Violence and incitement**, Health, Misinformation, Safety	"**Overturned** Facebook's decision to remove a post which it claimed, "contributes to the risk of imminent... physical harm." The Board found Facebook's misinformation and imminent harm rule (part of its Violence and Incitement Community Standard) to be inappropriately vague and recommended, among other things, that the company create a new Community Standard on health misinformation."
5. 2020-007-FB-FBR	12 Feb.	France, India, **Violence and incitement**, Religion, Violence	"**Overturned** Facebook's decision to remove a post under its Violence and Incitement Community Standard. While the company considered that the post contained a veiled threat, a majority of the Board believed it should be restored."
6. 2021-002-FB-UA	13 Apr.	Netherlands, **Hate speech**, Culture, Children / Children's rights, Photography	"**Upheld** Facebook's decision to remove specific content that violated the express prohibition on posting caricatures of Black people in the form of blackface, contained in its Hate Speech Community Standard."
7. 2021-003-FB-UA	29 Apr.	India, **Dangerous individuals and organizations**, Politics	"**Overturned** Facebook's decision to remove a post under its Dangerous Individuals and Organizations Community Standard. After the Board identified this case for review, Facebook restored the content. The Board expressed concerns that Facebook did not review the user's appeal against its original decision. The Board also urged the company to take action to avoid mistakes which silence the voices of religious minorities."

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

8. 2021-001-FB-FBR	5 May	United States, **Dangerous individuals and organizations**, Freedom of expression, Politics, Safety	"**Upheld** Facebook's decision on January 7, 2021, to restrict then-President Donald Trump's access to posting content on his Facebook page and Instagram account."
9. 2021-006-IG-UA	8 Jul.	United States, Turkey, **Dangerous individuals and organizations**, Freedom of expression, Marginalized communities, Misinformation	"**Overturned** Facebook's original decision to remove an Instagram post encouraging people to discuss the solitary confinement of Abdullah Öcalan, a founding member of the Kurdistan Workers' Party (PKK). After the user appealed and the Board selected the case for review, Facebook concluded that the content was removed in error and restored it. The Board is concerned that Facebook misplaced an internal policy exception for three years and that this may have led to many other posts being wrongly removed."
10. 2021-007-FB-UA	11 Aug.	Myanmar, **Hate speech**, Freedom of expression, Politics	"**Overturned** Facebook's decision to remove a post in Burmese under its Hate Speech Community Standard. The Board found that the post did not target Chinese people, but the Chinese state. Specifically, it used profanity to reference Chinese governmental policy in Hong Kong as part of a political discussion on the Chinese government's role in Myanmar."
11. 2021-008-FB-FBR	19 Aug.	Brazil, **Violence and incitement**, Governments, Health	"**Upheld** Facebook's decision to leave up a post by a state-level medical council in Brazil which claimed that lockdowns are ineffective and had been condemned by the World Health Organization (WHO). The Board found that Facebook's decision to keep the content on the platform was consistent with its content policies. The Board found that the content contained some inaccurate information which raises concerns considering the severity of the pandemic in Brazil and the council's status as a public institution. However, the Board found that the content did not create a risk of imminent harm and should, therefore, stay on the platform."
12. 2021-009-FB-UA	14 Sep.	Israel, Egypt, **Dangerous individuals and organizations**, News events, Journalism, War and conflict	"**Agrees that Facebook was correct** to reverse its original decision to remove content on Facebook that shared a news post about a threat of violence from the Izz al-Din al-Qassam Brigades, the military wing of the Palestinian group Hamas. Facebook originally removed the content under the Dangerous Individuals and Organizations Community Standard, and restored it after the Board selected this case for review. The Board concludes that removing the content did not

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011513

			reduce offline harm and restricted freedom of expression on an issue of public interest."
13. 2021-010-FB-UA	27 Sep.	Colombia, **Hate speech**, Community organizations, Freedom of expression, Protests	"**Overturned** Facebook's decision to remove a post showing a video of protesters in Colombia criticizing the country's president, Ivan Duque. In the video, the protesters use a word designated as a slur under Facebook's Hate Speech Community Standard. Assessing the public interest value of this content, the Board found that Facebook should have applied the newsworthiness allowance in this case."
14. 2021-011-FB-UA	28 Sep.	South Africa, **Hate speech**, Governments, Marginalized communities, Politics	"**Upheld** Facebook's decision to remove a post discussing South African society under its Hate Speech Community Standard. The Board found that the post contained a slur which, in the South African context, was degrading, excluding and harmful to the people it targeted.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_011514

Exhibit 288

rumble

August 31, 2022

TO: **British Columbia's Office of the Human Rights Commissioner**

ATTN: **Ms. Sarah Khan**
[redacted]@bchumanrights.ca

RE: **Rumble Inc.'s Response to Order to Produce Records**

Dear Ms. Khan:

Please find Rumble Inc.'s ("Rumble") response to British Columbia's Office of the Human Rights Commissioner's ("HRC") Order to Produce Records ("Order") dated May 31, 2022:[1]

1. N/A – Rumble does not in the ordinary course of business track the country or territory of origin of individual users. Rumble would have to make inquiries and perform additional analysis to generate location information for some of its users. This is not information that is currently compiled or organized in records in Rumble's custody or control.

2. Attached as Schedule "A" is Rumble's current Content Policies and Complaint Procedure (the "Content Policies"). Attached as Schedule "B" is a press release dated June 16, 2022, describing Rumble's proposed future Rules for Prohibited Content and Removal and Appeal Process (the "Rumble Rules").

3. N/A – As indicated, Rumble does not track individual users by location and is therefore unable to provide information responsive to this request. In addition, although enforcement of Rumble's Content Policies may result in the removal of offensive content and/or users, Rumble does not regularly compile records tracking such incidents.

4. Rumble's Content Policies guide the work of Rumble's content moderators. The Content Policies have not changed since January 2020. The Rumble Rules—Rumble's proposed future rules regarding content moderation—are attached as Schedule "B".

5. N/A – Rumble does not define "borderline" content. Moderation of content is based upon review and assessment against the Content Policies.

6. Content moderation is conducted based upon review and removal of content prohibited by Rumble's Content Policies. Rumble's plan is to implement automated processes in the future for the identification and removal of copyright infringing and pornographic material as outlined in the

[1] Following receipt of the Order, the HRC and Rumble agreed to extend the deadline for the delivery of this response until August 31, 2022. Rumble provides this response in order to assist the HRC and without prejudice to any position it may take in response to future inquiries by the HRC.

Rumble Rules. No internal studies or reports of the sort requested are in Rumble's custody or control.

7. Rumble's current processes for identifying and acting upon content that violates Rumble's Content Policies consist of (i) independent content moderators flagging content that violates Rumble's Content Policies; and (ii) flagging of content by other users of Rumble pursuant to the Complaint Procedure outlined in the Content Policies. Rumble's future anticipated processes for addressing content that violates the Rumble Rules is detailed in Schedule "B". As noted, Rumble does not track individual users by location and does not keep a running log of content and/or user removals.

8. No.

9. No, Rumble does not have content moderators specifically for British Columbia.

10. Rumble does not evaluate the risk of algorithms as they relate to content that would violate Rumble's Content Policies and therefore has no internal research in relation thereto.

11. Rumble has provided information to the Organization for Economic Co-operation and Development ("OECD") in response to the OECD's project on online dissemination of terrorist and violent extremist content ("TVEC"). Information gathered by the OECD will be used to further its Voluntary Transparency Reporting Framework initiative for TVEC.

rumble

Schedule A
Current Content Policies and Complaint Procedure

(Also available at https://rumble.com/s/terms)

RUMBLE CONTENT POLICIES

Rumble enables creators of Content to submit Content to the Rumble Service in accordance with the Agency Options set forth below. Under no circumstances shall the following be submitted to the Rumble Service:

1. Content or material which was created by a person other than you, or subject to copyright by a person other than you, without that person's express written authorization;
2. Content which you have distributed to any person by any means other than via the Rumble Service;
3. Content or material which includes, in whole or in part, audio, video, text, or images that are subject to copyright by another person unless such inclusion is subject to applicable laws concerning fair use or is otherwise expressly authorized by the copyright holder;
4. Content or material that is subject to any third-party contractual provisions or limitations, including but not limited to license restrictions;
5. Content or material which displays the trademarks of another person without that person's authorization, subject to applicable laws concerning fair use;
6. Content or material that is pornographic, obscene, or of an adult or sexual nature;
7. Content or material that is grossly offensive to the online community, including but not limited to, racism, anti-semitism and hatred;
8. Content or material that;
    . Promotes, supports, or incites violence or unlawful acts;
    a. Promotes, supports or incites individuals and/or groups which engage in violence or unlawful acts, including but not limited to Antifa groups and persons affiliated with Antifa, the KKK and white supremacist groups and or persons affiliated with these groups; and/or
    b. Promotes or supports entities and/or persons designated by either the Canadian or United States government as terrorists or terrorist organizations.
9. Content or material that exploits children under the age of 18 or posts or discloses any personally identifying information about any person at any age, including but not limited to personally identifying information about children under the age of 18;
10. Content or material promoting or providing instructional information about illegal activities, promoting harm or injury to any group, individual or cruelty to animals including, but not limited to:
    . Disseminating personal information about another individual for malevolent purposes, including libel, slander, "doxxing", defamation or violation of an individual's right to privacy.
11. Content or material that infringes or encroaches on the rights of others, including, but not limited to, infringement of privacy, publicity rights, and harm to reputation;
12. Any Content or material that contains links to another website where any of the above-described prohibited Content or material is accessible.

CAN_DMS: \147617815

CONFIDENTIAL TREATMENT REQUESTED RUMBLE_HJC_00000010

13. Any other Content or material that Rumble in its sole, unfettered, and arbitrary discretion, determines is undesirable on the Rumble Service.

Rumble has the sole discretion to decide whether Content or material is permitted on the Rumble Service and any materials submitted to the Rumble Service may be, but is not necessarily, examined by Rumble before it is made available on the Rumble Service. You acknowledge that Rumble has the absolute right (but not the obligation) to prohibit, refuse, delete, move and edit Content and material for any reason, in any manner, at any time, without notice to you.

You should also know that in visiting the Rumble Site or viewing Content via the Rumble Player or otherwise, you may be exposed to materials which you consider to be offensive or inappropriate and you assume the risk and sole responsibility for your exposure to any such Content or material.

To report content that violates our policies, please email moderation@rumble.com

COMPLAINT PROCEDURE

We take seriously any complaints received from viewers of Rumble Content and attempt to respond to every complaint. If you have a complaint regarding the Content, materials available through the Rumble Service, inappropriate behavior or postings by other users in the Forum, or otherwise, you may submit your complaint to Rumble by emailing our customer service representatives at support@rumble.com. Please use the word "Complaint" in the subject line, and include the following information in your email:

1. your name, email address and, if you are a registered user, your user ID;
2. the basis of your complaint (provide as much detail as possible);
3. if your complaint concerns the activities of other users/visitors on the Forum identify the specific type of inappropriate or offensive behavior engaged in and, insofar as possible, the identity of the offending person. If your complaint concerns particular Content please provide the reason for your Complaint and the title and/or location of any video; and
4. the date(s) on which you observed the objectionable activities or behaviour or the date on which you viewed the Content which is the subject of your complaint;

A customer service representative will endeavour to respond to your email and if, in Rumble's determination, your complaint is a valid one, Rumble will take appropriate actions in its sole discretion, and has no responsibility to at any time report to you as to the status or outcome of its investigation or any actions Rumble has taken as a result.

Schedule B
Future Content Policy
Press Release & Proposed Rumble Rules

(Also available at https://corp.rumble.com/blog/rumble-proposes-an-open-source-content-moderation-policy-process-to-improve-transparency-put-creators-first/)

First phase of a content creator-designed moderation process will help the video sharing platform stay true to its mission as it continues to grow; changes will be incorporated into terms of use after feedback from creators and users.

Longboat Key, Fla., June 16, 2022 – Today, the video streaming platform Rumble announced the first phase of an updated content moderation process designed by creators and anchored in transparency. The proposed moderation policies were designed by leading Rumble and Locals creators Robert Barnes and David Freiheit, both of whom are also accomplished attorneys. Under the proposed policies, content creators will be able to express themselves to interested audiences within the limits of the law and without harassment while ensuring a consistent and transparent process as the platform continues its rapid growth. Rumble welcomes feedback on the proposed policies from the platform's creators and users, which it will consider before implementing the new policies as part of the platform's terms of use later this year.

"Despite the mainstream pressure, we have stayed true to our original content moderation philosophy since Rumble was founded in 2013," said CEO Chris Pavlovski. "Our mission is to protect a free and open internet, and we will never 'move the goal posts' on our community unless required by law. We're putting creators and users first by asking their views on the policies before we implement them. It's a philosophy that we feel is incredibly important for our society and one that has progressively differentiated our company from big tech platforms that continue to embrace 'cancel culture'."

"Unlike other tech platforms, Rumble's proposal will provide a free space for open discourse without politicized discrimination in the digital public square, while simultaneously protecting users from harassing behavior," said Robert Barnes, founder of Barnes Law LLP and the creator of the "Barnes Law" video channel on Rumble.

"We seek clear, transparent rules and non-discriminatory enforcement without political or partisan bias." said David Freiheit, founder of Freiheit Legal and the creator of the "Viva Frei" video channel on Rumble. "By starting with input from creators, these new content policies will help Rumble hold true to its mission of neutrality and ensure transparency for creators and users alike."

CAN_DMS: \147617815

CONFIDENTIAL TREATMENT REQUESTED

Today, the first phase of the new content moderation process, Rumble released a new set of proposed content policies and removal and appeals process. [See the proposed "Rumble Rules" and removal and appeals process further below and provide feedback to rumblerules@rumble.com].

In the next phase, Rumble will seek feedback from its creators and users on the policies and establish a new appeals process that will be adjudicated by a community of Rumble creators. To ensure transparency, Rumble will ensure that all appeal decisions are made readily available to the public. The full implementation of this process, including the incorporation of the new content policies and the appeals process in the platform's terms of use, is expected to be completed by the end of the year.

Rumble is a high-growth neutral video platform that is creating the rails and independent infrastructure designed to be immune to cancel culture. Rumble's mission is to restore the Internet to its roots by making it free and open once again. Additionally, the company announced in December 2021 the execution of a definitive business combination agreement with CF Acquisition Corp. VI (NASDAQ: CFVI). See the announcement here: https://corp.rumble.com

Proposed Rumble Rules: Prohibited Content

Rumble seeks to provide an open space for content creators to freely express themselves to an engaged audience, within the limits of the law of their locale and without illicit misuse of the platform that could undesirably discourage open debate and free discussion of publicly important topics. In an era of cancel culture and de-platforming, Rumble is, and shall remain, a free space for open discourse. We aim to simultaneously protect users from illicit stalking behavior by bad faith users of the platform whose behavior discourages use of the service and who engage in unlawful discrimination. We provide these clear, transparent rules, and non-discriminatory enforcement thereof, without political or partisan bias. These rules are provided free of copyright and are opensource to all to copy and replicate in this effort to revolutionize the space for free speech on the internet and in the digital public square.

The freedom of expression Rumble seeks to preserve and protect on its platform protects against unlawful discriminatory use and, equally, prohibits discriminatory abuse of the service. It is the freedom to freely express oneself on subjects that may run contrary to the official narrative without fear of sanction or censorship from the platform. It is the freedom to hold and express unpopular beliefs. It is not the freedom to break the law or stalk others or misuse the platform to target people based on

legally protected status. Rumble is not here to suppress the freedom of speech of others, nor to stalk or suppress others' expression.

In that spirit, these are the proposed Rumble Rules:

All content uploaded to Rumble must abide by the applicable laws in your jurisdiction. Prohibited use of this service includes:

a. using the service to incite or invite unlawful conduct, such as infringing legally-protected trademarks, unauthorized publication of legally-protected copyrighted material, confidential information, trade secrets, or proprietary information, and other such illicit acts, such as inciting violence, defaming others, doxing others, or illicitly invading the privacy of another individual;
b. publishing obscenity, including pornography and nudity for a prurient purpose;
c. stalking, such as misusing the platform to make others uncomfortable in their use of the platform through a pattern of continuous conduct, creating a reasonable fear in another person for their personal safety or deterring a reasonable user from use of the service in fear of such digital stalking;
d. discriminating against others based on their legally protected status, such as attacking other users or content creators on the platform, based on that user's race, religion, or other legally protected status;
e. unauthorized advertising, soliciting information from a person under the age of 18, harvesting or collecting information for the purpose of unsolicited bulk communications, or using the service for any fraudulent purpose.

We will ensure that we comply with all applicable laws, while providing the most robust and civil platform for creative expression and meaningful discourse. Less censorious overseers and less hateful trolls. That is our promise to our content creators, our viewers, and the entire world.

Rumble on.

Rumble Removal & Appeal Process

Rumble's commitment to our community – of both content creators ("**Creator(s)**"), and content consumers ("**Consumer(s)**") – is to ensure clarity, transparency, and the impartial and apolitical application of our rules and terms of service. While the free and meaningful expression of ideas is the ethos of our proverbial digital public square, we have established certain rules to remain compliant with applicable laws, and to foster and protect an environment in which the free and meaningful expression of ideas can in fact occur. We strive to implement these rules fairly and transparently,

but it is possible that a Creator believes a video has been wrongly removed from the platform. This is our removal and appeals process:

Automated flagging

Rumble will not employ automated flagging of uploaded content with the exception of copyright infringement and pornographic material ("**Automated Flagging**").

Manual Review

With the exception of Automated Flagging, all other identification of potentially contravening content ("**Alleged Contravening Content**") is done on an ad hoc basis by Creators or Consumers ("**Community Identification**"). Because Community Identification is both a right and a privilege, any Creator or Consumer who engages in wrongful, abusive, and/or unjustified Community Identification of content may be sanctioned for such behavior as provided for herein.

Community Identification

In the event that a Creator or Consumer flags or otherwise reports Alleged Contravening Content, such content will be subject to manual review by Rumble representatives ("**Rumble Rep**"). In the event that the Rumble Rep determines that the Community Identification was ill-founded and does not lead to any further action by Rumble (an "**Unsubstantiated Complaint(s)**"), the Creator or Consumer shall receive a warning for having submitted an Unsubstantiated Complaint.

Sanction for Unsubstantiated Complaints

A Creator or Consumer who submits three (3) Unsubstantiated Complaints within a four (4) month period shall receive a first warning, which shall result in the suspension of Rumble services to that Creator or Consumer for one (1) week (a "**First Strike**"). A Creator or Consumer who submits an additional three (3) Unsubstantiated Complaints within a subsequent four (4) month period shall be subject to suspension of two (2) weeks (a "**Second Strike**"). And a Creator or Consumer who submits an additional three (3) Unsubstantiated Complaints within a subsequent four (4) month period shall be blocked from Rumble for a period of one (1) year.

Copyright Infringement Complaints

In case of alleged copyright infringement, Rumble will immediately de-monetize the allegedly infringing content and provide notice to the Creator. If the Creator removes

the allegedly infringing content from Rumble or provides evidence of authorized use within twelve (12) hours of notification, at the sole discretion of Rumble, no strike shall be assessed against the Creator.

Rumble Removal Notice & Sanctions

In the event that a Rumble Rep deems the Community Identification to be legitimate or otherwise substantiated, Rumble shall remove the Contravening Content from the platform without further notice or delay, save and except for a notification via email to the Creator who uploaded the Contravening Content advising them of the removal (a "**Rumble Removal Notice**").

A Creator who receives three (3) Rumble Removal Notices within six (6) months of a first Rumble Removal Notice shall have their Rumble account terminated. In the event that a Creator who has received a Rumble Removal Notice does not receive three (3) such notices within six (6) months of the issuance of the first such Rumble Removal Notice, such notices shall automatically extinguish six (6) months after the issuance of the any such Rumble Removal Notice. Other than overtly illegal content or violent content, as determined at the sole discretion of Rumble, all terminations will be for a period of one (1) year, after which the Creator may apply for the restoration for their account.

Immediate Termination

If a Creator receives a Rumble Removal Notice for content that, in the opinion of the Rumble Rep, is sufficiently egregious or clearly in violation of any applicable law, Rumble may terminate the account immediately, subject to appeal as provided below.

Appeal to the Rumble Community

In the event that a Rumble Rep issues a Rumble Removal Notice to a Creator, and that Creator believes the Rumble Removal Notice was unwarranted, illegitimate and unjustified (an "**Illegitimate Removal Notice**"), that Creator shall have the right, within a period of fifteen (15) days of the issuance of the Rumble Removal Notice, to appeal the issuance of the Rumble Removal Notice to the "**Rumble Quality Control Community Members**" ("**RQCCM**").

Composition of the RQCCM

The RQCCM shall consist of Rumble Creators with a proven track record of consistent and active content creation, who are in good standing within the Rumble community, and have a sufficient following on the platform, as determined by

Rumble. The criteria for RQCCM shall be disclosed to the Rumble community and may be updated or modified from time to time.

The RQCCM panel shall consist of no less than nine (9) Creators and shall always consist of an uneven number of Creators so as to avoid a potential tie on an RQCCM review. The RQCCM shall, within seventy-two hours, by simple majority, have the final determination in affirming or reversing the initial decision of the Rumble Rep on the Rumble Removal Notice. The results of the RQCCM review, and any explanatory notes as the case may be, shall be made available to the Creator and the Rumble community for full transparency.

The identities of the specific RQCCM on any given decision shall not be publicly disclosed. However, a list of the Creators who are RQCCM shall be made public and updated from time to time, as required. RQCCM may receive remuneration of preferential CPM premiums for their participation in the RQCCM service they render for the Rumble Community in the context of this Rumble appeal process. In the event that the RQCCM reverse a Rumble Removal Notice that had been issued by a Rumble Rep, all the rules provided for herein as relates to Unsubstantiated Complaints shall apply *mutatis mutandis.* For the avoidance of doubt, and notwithstanding anything contained herein to the contrary, this removal and appeals process is in addition to, and does not replace, any applicable notification or counter-notification actions or obligations contained in the Digital Millennium Copyright Act ("DMCA"). Rumble's DMCA procedures are outlined at rumble.com.

CAN_DMS: \147617815

Exhibit 289

ANNEX : Information on the risk assessment reports

Expectations

As explained in recital 100 of Regulation (EU) 2022/2065, you should **report comprehensively on the risk assessments performed**. In this regard, risk assessment reports should provide a comprehensive presentation at least of the following elements:

1. The systemic risks in the Union identified, analysed and assessed, stemming from the design or functioning of the service and its related systems, including algorithmic systems, or from the use made of the service – pursuant to Article 34(1) of Regulation (EU) 2022/2023. The report should elaborate on each of the categories of risks pursuant to that article in a way that is specific to the service and the risk profile of the service, and proportionate to the systemic risks, taking into account their severity and probability, notably as regards:
    a. the dissemination of illegal content through their services;
    b. any actual or foreseeable negative effects for the exercise of fundamental rights, in particular the fundamental rights to human dignity enshrined in Article 1 of the Charter, to respect for private and family life enshrined in Article 7 of the Charter, to the protection of personal data enshrined in Article 8 of the Charter, to freedom of expression and information, including the freedom and pluralism of the media, enshrined in Article 11 of the Charter, to non-discrimination enshrined in Article 21 of the Charter, to respect for the rights of the child enshrined in Article 24 of the Charter and to a high-level of consumer protection enshrined in Article 38 of the Charter;
    c. any actual or foreseeable negative effects on civic discourse and electoral processes, and public security;
    d. any actual or foreseeable negative effects in relation to gender-based violence, the protection of public health and minors and serious negative consequences to the person's physical and mental well-being.

2. The assessment should be detailed per risk factors identified, in accordance with Article 34(2) of Regulation (EU) 2022/2065. In particular, the reports should elaborate on the assessments made with respect to relevant algorithmic systems such as recommender systems, content moderation systems and applicable terms and conditions and their enforcement, advertising systems, data related practices. The assessment should not be limited only to the design of the systems, but also to their functioning and use. For example, an important risk factor for analysis refers to intentional manipulation of the service, including inauthentic or automated exploitation of the service and the amplification of illegal content – such as dangerous goods – or information that is incompatible with terms and conditions – such as content restricted to minors, or proven misinformation.

3. Throughout the report, it should be clear how the risk assessment took into account specific regional or linguistic aspects, including when specific to a Member State. For example, in view of the prominence and severity of risks in electoral periods, the assessments of the relevant services should clarify how the national and European elections are considered with respect to the providers' obligations

Commission européenne/Europese Commissie, 1049 Bruxelles/Brussel, BELGIQUE/BELGIË – Tel. Redacted
Office: L-51 04/DCS – Tel. direct line Redacted

CNECT-DSA-REGISTRY@ec.europa.eu

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

regarding risks for electoral processes and civic discourse. For other service providers, the risk profile of their services could be rather linked to the dissemination of certain illegal goods, potentially in specific periods of the year.

4. Diligence in the identification, analysis and assessment of the risks specific to the service could be illustrated, among other, for example by explaining why the provider considers the assessment proportionate to the risks identified, which risks were not retained for an in-depth assessment and why, or how they ensured that the risk assessments are informed by the 'best available information and scientific insights' and how they tested 'their assumptions with the groups most impacted by the risks', as detailed in Recital 90 of Regulation 2022/2065. They should also point to the evidence underpinning the assessments and the methodologies used.

Providers should take note that risk mitigation obligations pursuant to Article 35 of Regulation (EU) 2022/2065 should be tailored to the specific risks identified pursuant to Article 34 of that Regulation. Such measures should be reasonable, proportionate and effective to mitigate those risks. Article 42(4) of that Regulation also requires that providers report to the Commission on those specific mitigation measures, without undue delay upon completion.

Consequently, it is important that the risk assessment reports focus on the identification, analysis and assessment of systemic risks, whereas reports on risk mitigation elaborate on those measures.

Providers are also reminded that they must preserve any documents supporting their risk assessments for at least three years, in accordance with Article 34(3) of Regulation (EU) 2022/2065.

Other considerations

In accordance with Article 41 of Regulation (EU) 2022/2065, the compliance function has a particularly important role to play in ensuring that the risks referred to in Article 34 of that Regulation are identified and properly reported on and that reasonable and effective risk-mitigation measures are taken.

To this end, it is of paramount importance that the compliance function has sufficient authority within the organisation and has stature, for example to ensure smooth interactions with the product, integrity, legal or trust and safety teams involved in ensuring compliance with the risk assessment obligations. To this end, the compliance function should be appropriately resourced, both in terms of staff numbers and their competences, and have access to information and relevant systems.

The head of the compliance function should report directly to the management body of the provider, pursuant to Article 41(2), third sub-paragraph. This provision is particularly important also in light of the obligations the management body bears in accordance with Article 41 paragraphs 5 to 6. The management body should be actively involved in the sound management of the systemic risks identified pursuant to Article 34. The management body shall approve and review periodically, at least once a year, the strategies and policies for taking up, managing and mitigating risks. It shall also devote sufficient time and resources to the consideration of the measures related to risk management. It shall be actively involved in the decisions related to risk management and bears the responsibility for ensuring that adequate resources are allocated to the management of risks.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_007010

In view of these requirements, risk assessments should not constitute a mere reporting obligation for the provider, but should be regarded as a structuring tool in the design, functioning and use of the service. It should be embedded in the top management decisions.

Additional resources

The European Commission organised on 27 June this year a stakeholder event focusing on the priority areas for the implementation of the Digital Services Act. During this event, panellists expressed their views on how risk assessments should be conducted by very large online platforms and search engines. These discussions can be viewed here and a synthesis report will be made available at the same online address shortly.

On 9th June the European Commission hosted a round-table with civil society organisations, who raised substantive expectations as to what risk assessments should entail. A Chatham House report from the discussion will be available here.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_007011

Exhibit 290

Ref. Ares(2023)7112552 - 19/10/2023

EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR COMMUNICATIONS NETWORKS, CONTENT AND TECHNOLOGY

Platforms Policy and Enforcement
Digital Services

Brussels, 19 October 2023
SENSITIVE (*)

TikTok Technology Limited
10 Earlsfort Terrace
Dublin
D02 T380
Ireland

Redacted via email:
Redacted tiktok.com

Subject: Request for information pursuant to Article 67(1) and Article 34(3) of Regulation (EU) 2022/2065 regarding TikTok's compliance with Regulation (EU) 2022/2065, in particular with its provisions 16(6), 34, and 35.

Dear Madam,

On 25 April 2023, TikTok was designated as a very large online platform ('VLOP') pursuant to Article 33(4) of Regulation (EU) 2022/2065 of the European Parliament and the Council of 19 October 2022 on a Single Market for Digital Service (the Digital Services Act). ([1]) On the following day, that designation decision was notified to TikTok, and an acknowledgement of receipt was received by the Commission services on 27 April 2023. Consequently, pursuant to Articles 33(6) and 92 of Regulation (EU) 2022/2065, that regulation started to apply to TikTok as from 28 August 2023.

Pursuant to Article 16 of Regulation (EU) 2022/2065, providers of VLOPs are obliged to put mechanisms in place to allow any individual or entity to notify them of the presence on their service of specific items of information that the individual or entity considers to be illegal content.

(*) Distribution only on a need-to-know basis. Do not read or carry openly in public places. Must be encrypted in transmission and stored securely, where possible by using encrypted storage. Destroy copies by shredding or secure deletion. Full handling instructions: https://europa.eu/!db43PX

([1]) OJ L 277, 27.10.2022, p. 1–102.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Providers of VLOPs are obliged to process notices that they receive under such mechanisms and take their decisions in respect of the information to which the notices relate, in a timely, diligent, non-arbitrary and objective manner.

Pursuant to Article 34 of Regulation (EU) 2022/2065, providers of VLOPs are obliged to diligently identify, analyse, and assess any systemic risks in the Union stemming from the design or functioning of their service and its related systems, or from the use made of their services. Article 34(1)(a),(b) and (d) list risks related to the dissemination of illegal content and to actual or foreseeable negative effects for the exercise of the rights of the child, as enshrined in Article 24 of the European Charter of Fundamental Rights, and on the protection of minors as well as serious consequences on the person's physical and mental well-being. Article 34(1)(c) lists risks related to actual or foreseeable negative effects on civic discourse and electoral processes as one of the risks to be assessed in that risk assessment. Pursuant to Article 34(2), third subparagraph, risk assessments must take into account specific regional or linguistic aspects, including in relation to a Member State.

Pursuant to Article 34(3) of Regulation (EU) 2022/2065, providers of VLOPs are required to preserve the supporting documents of the risk assessment for at least three years after the performance of risk assessments, and upon request, communicate them to the Commission.

Pursuant to Article 35 of Regulation (EU) 2022/2065, providers of VLOPs are obliged to put in place reasonable, proportionate and effective mitigation measures, tailored to the specific systemic risks identified, including in relation to the dissemination of illegal content, any actual or foreseeable negative effects for the exercise of the rights of the child, on civic discourse and electoral processes and public security, the protection of minors and the serious negative consequences on the person's physical and mental well-being.

Pursuant to Article 42(4) of Regulation (EU) 2022/2065, providers of VLOPs are obliged to transmit their risk assessment reports to the Commission without undue delay upon completion. On 29 September, TikTok submitted its risk assessment report to the Commission. The report also includes information about mitigation measures under Article 35(1) of Regulation (EU) 2022/2065.

Hamas - Israeli conflict

On 7 October 2023, Hamas, an organisation subject to restrictive measures pursuant to Council Implementing Regulation (EU) 2022/147 and recognised as a terrorist organisation by the Council Common Position of 27 December 2001 on the application of specific measures to combat terrorism (2001/931/CFSP) ([2]), launched a terror attack against the state of Israel.

By letter issued on 12 October 2023, Commissioner Breton reminded TikTok of its obligations regarding content moderation laid out in Regulation (EU) 2022/2065 with particular regard to illegal content and disinformation disseminated in the context of the Hamas-Israeli conflict and of

([2]) eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32001E0931&qid=1412597066958&from=EN

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_006392

the challenges in tackling disinformation with respect to the recent and forthcoming elections in the EU.

On 13 October 2023, Mr Shou Chew, Chief Executive Officer of TikTok, replied by letter outlining the actions taken by TikTok as regards the Hamas-Israeli conflict.

On 15 October 2023, TikTok published a news item ([3]) outlining the actions taken to protect the TikTok community during the Israel-Hamas conflict.

However, based on the information contained in TikTok's reply to Commissioner Breton's letter of 13 October 2023[4], TikTok's news item published on 15 October 2023 and TikTok's risk assessment report, the Commission is unable to assess whether TikTok conducted its risk assessment diligently with respect to the systemic risks listed in Article 34(1), as detailed in Annex I. The Commission is also unable to assess whether the risk mitigation measures put in place by TikTok, pursuant to Article 35, are reasonable, proportionate and effective.

Elections

By letter of 5 September 2023 (Ref: Ares(2023)6026012) the Commission services asked TikTok several questions related to the civic discourse and electoral processes systemic risks which TikTok, as a VLOP, must assess pursuant to Article 34(1) of Regulation (EU) 2022/2065. On 8 September 2023, TikTok submitted written input about its risk mitigation measures in relation to the forthcoming Slovak elections.

On 12 September 2023, the Commission services held a meeting with TikTok in Bratislava during which TikTok's risk mitigation measures in relation to the forthcoming Slovak elections were discussed. The Commission services also take note of TikTok's policies communicated in the risk assessment report and during the meeting of 12 September 2023 on 'Election Misinformation, 'Synthetic and Manipulated Media', and policies related to hate speech, used as measures mitigating the risks of negative effects on civic discourse and electoral processes.

On 30 September 2023, parliamentary elections were held in Slovakia. Those elections were the first elections held in the EU after Regulation (EU) 2022/2065 started to apply to TikTok as a VLOP. On 8 October 2023 general elections were held in Luxembourg. On 15 October 2023, parliamentary elections were held in Poland.

Based on publicly available information ([5]), the Commission has become aware of content circulated on TikTok in the lead up to the Slovak parliamentary elections which appear to violate

---

([3]) Our continued actions to protect the TikTok community during the Israel-Hamas war | TikTok Newsroom

([4]) The Commission took note of TikTok's request to refrain from registering TikTok's reply of 13 October 2023 in TikTok's file pending a discussion with Commission services.

([5]) Trust Lab, "Code of Practice on Disinformation: A Comparative Analysis of the Prevalence and Sources of Disinformation across Major Social Media Platforms in Poland, Slovakia, and Spain", Semi-Annual Report, September 2023; European Commission, Directorate-General for Communications Networks, Content and

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TikTok's policies outlined above and that may appear to contribute to risks linked to civic discourse and electoral processes.

Other aspects

Based on information provided in TikTok's risk assessment report, the Commission is unable to assess whether the systemic risks specific to TikTok and in particular vis-à-vis minors, have been diligently identified, analysed and assessed and whether the relevant mitigation measures put in place are reasonable, proportionate and effective.

The questions included in this letter are without prejudice to additional questions the Commission may address to TikTok, also in particular for the obligations related to the protection of minors under Regulation (EU) 2022/2065.

Pursuant to Article 67 of Regulation (EU) 2022/2065 the Commission may, by simple request, require a provider of a VLOP to provide information on a suspected infringement of that regulation. Pursuant to Article 65 (1), the Commission may make such a request before initiating formal proceedings.

Pursuant to Article 34(3) of Regulation (EU) 2022/2065, the Commission may further request providers of VLOPs to submit certain supporting documentation on which they relied upon to draw up the risk assessment report.

On these bases, the Commission requests TikTok to provide the information listed in the questionnaire in Annex I of this letter which will enable the Commission to further assess compliance with Articles 16(6), 34 and 35 of Regulation (EU) 2022/2065.

i. Concerning the requests under section I (TikTok's application of its obligations in the context of the Hamas-Israeli conflict – compliance with Articles 16(6), 34 and 35(1) of Regulation (EU) 2022/2065), please reply no later than **25/10/2023**;
ii. Concerning the requests under sections II and III (TikTok's risk assessment and risk mitigation measures in relation to civic discourse and electoral process and measures related to minors – compliance with Articles 34 and 35(1) of Regulation (EU) 2022/2065), please reply no later than **8/11/2023**.

Please submit your reply by sending it to the functional mailbox of the Directorate General for Communications Networks, Content and Technology at CNECT-DSA-Registry@ec.europa.eu

If you request that any piece of information supplied in your response should be treated as confidential and should not be published in any public version of Commission's documents in the future related to TikTok's compliance with the Regulation (EU) 2022/2065, please submit both a confidential and a non-confidential version of your reply and provide reasons for any

Technology, "Digital Services Act- Application of the risk management framework to Russian disinformation campaigns", Publications Office of the European Union, 2023.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_006394

confidentialities claimed. Unless you specify otherwise, your response shall be treated as non-confidential.

Any personal data submitted in your reply will be processed in compliance with Regulation (EU) No 2018/1725 of the European Parliament and of the Council. ([6])

If you consider that some documents included in your reply (or parts of some documents) contain personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership or genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation ([7])(i.e. special categories of personal data), please provide those documents in a separate submission identified as "SPECIAL CATEGORIES OF PERSONAL DATA", linking each document to the question number to which the document constitutes a reply. Please highlight the special categories of personal data in the document and provide an explanation as to the specific category in case this is not immediately obvious from the information highlighted.

This communication and its Annexes constitute a simple request for information pursuant to Article 67(1) of Regulation (EU) 2022/2065, which empowers the Commission to require providers provide all necessary information to enable it to carry out its duties under that Regulation.

Please note that, pursuant to Article 74(2)(a) of Regulation (EU) 2022/2065, the Commission may impose on a provider fines not exceeding 1 % of their total worldwide turnover in the preceding financial year where they intentionally or negligently supply incorrect, incomplete or misleading information in response to a simple request for information pursuant to Article 67.

If you have any further questions, please address them by e-mail to CNECT-DSA-Registry@ec.europa.eu, clearly indicating the subject of this letter and the reference number.

([6]) Regulation (EU) 2018/1725 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC (OJ L 295, 21.11.2018, p. 39).

([7]) See Article 10(1) of Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data (EU-DPR), and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC (OJ L 295, 21.11.2018, p.39), which is identical to Article 9(1) of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (GDPR) (OJ L119, 4.5.2016, p.1).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_006395

Yours faithfully,

Redacted

- Annex I: Questionnaire
- Annex II: Instructions for replying

c.c.: Redacted tiktok.com; Redacted bytedance.com;
Redacted tiktok.com; Redacted tiktok.com

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_006396

ANNEX I

QUESTIONNAIRE
TikTok
DEADLINE FOR REPLY:
For Section I: 25/10
For Sections II and III: 8/11

1. BUSINESS SECRETS AND OTHER CONFIDENTIAL INFORMATION

Does your reply contain any parts that you regard as business secrets or other confidential information that should not be published or otherwise disclosed to third parties?

YES ☐ NO ☐

If yes, please submit both a confidential and a non-confidential version of your reply and provide reasons for any confidentialities claimed.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_006397

2. QUESTIONS

Format of the response

See Annex II.

Specific questions

In replying to the questions, please refer to measures as from 28 August 2023, unless otherwise specified in each question. Please do not limit your reply to the situation at the time of submitting your response but provide details on any changes which may have occurred since 28 August 2023 or the date otherwise specified in the specific question.

In several questions you are requested to provide information on performance metrics and internal controls:

- *By performance metrics, the Commission services mean in particular, but not only, quantitative performance indicators that help determine objectively the performance of a given policy, action or measure, or that allow the determination of progress towards reaching the objectives of the measure.*
- *By internal controls, the Commission services mean any type of measure, including business process and tests conducted and organisation arrangement, that aims to monitor and ensure the effectiveness and correct functioning of the measures put in place.*

Where your organisation is using different terminology, please explain the relevant organisational set-ups that are closest to the requested concepts.

I. As regards the obligations stemming from Regulation (EU) 2022/2065 in the context of the Hamas-Israeli conflict (compliance with Article 16(6), 34 and 35(1) of Regulation (EU) 2022/2065)

In the news item published on 15 October 2023, TikTok lists its "*continued actions to protect the TikTok community during the Israel-Hamas war*" (hereinafter referred as "TikTok's statement on continued actions to protect the TikTok community during the Israel-Hamas conflict") ([8]).

In responding to the questions under this subsection, please do not limit your reply to the situation at the time of submitting your response, but provide details on any changes which you may have implemented since 7 October 2023 (e.g., any change in organisation set up or increase/decrease in resources from the 7 October 2023 to the date of your response to this request for information).

([8]) Our continued actions to protect the TikTok community during the Israel-Hamas war | TikTok Newsroom

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

A. Questions related to "crisis management process" ([9]) put in place in the context of the Hamas-Israel conflict.

1. In TikTok's response of 13 October to Commissioner's Breton's letter of 12 October 2023, it is stated that TikTok engaged its "*crisis situation protocols*" immediately when it learned of the attack by Hamas, with 24/7 control rooms being established to monitor and address the emerging crisis. Please provide information and underlying internal documents related to the following questions:

    i. When was the "*crisis management process*"' (or crisis protocols) activated and for which duration (if already determined)?

    ii. What are the activation criteria for this level of crisis response?

    iii. Are there any geographical differences in the implementation of the "*crisis management process*"?

    iv. What is the interplay between TikTok's operations to address the crisis and the GIFCT Crisis Incident protocol in the context of the Hamas-Israeli conflict?

    v. What are the criteria for de-escalation of the crisis response?

    vi. Have you foreseen post-event review processes?

2. In its statement on the continued actions to protect the TikTok community during the Israel-Hamas conflict, TikTok states that it "*immediately mobilized significant resources and personnel to help maintain the safety of our community and integrity of our platform.*" TikTok also states that it launched a "*command center that brings together key members of our 40,000-strong global team of safety professionals, representing a range of expertise and regional perspectives, so that we remain agile in how we take action to respond to this fast-evolving crisis*". Please provide information and underlying documents related to the following questions:

    i. What are the concrete missions entrusted to the "*command center*"?

    ii. Please explain the organisational set up of the "*command center*" dealing with the current crisis, including organisation charts, reporting lines, information on the teams involved, their size and location.

    iii. The criteria for de-escalation of the "*command center*" element of the crisis response.

3. In its statement on the continued actions to protect the TikTok community during the Israel-Hamas conflict, TikTok states that it is adding "*more moderators who speak Arabic and Hebrew to review content related to these events*".

([9]) Or "crisis situation protocols", as you refer to it in the letter of 13 October.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_006399

i. Please indicate the current numbers of moderators who respectively speak Arabic and Hebrew, as well as the changes in number that have occurred since the beginning of the conflict.

ii. Please indicate the qualifications of the staff involved in content moderation linked to the current crisis, including their subject matter expertise.

B. Questions related to the applicable terms and conditions and their enforcement.

4. In its statement on the continued actions to protect the TikTok community during the Israel-Hamas conflict, TikTok states that it has adapted its "*proactive automated detection systems in real-time as [it] identif[ies] new threats; this enables [TikTok] to automatically detect and remove graphic and violent content so that neither [TikTok] moderators nor [TikTok] community members are exposed to it*".

i. Please describe how these automated detection systems are adapted "*in real time*".

ii. Please provide details of any measurement or assessment of accuracy of the enforcement of TikTok policy – i.e., the effectiveness of TikTok's measures - which covers, explicitly or implicitly, the management of the crisis as of 7 October 2023, and notably the impact in the European Union.

iii. Please provide any evidence including in the form of internal documents regarding tests of the adaptation of the automated detection systems, their outcome and the conclusions drawn from the tests, including the timeline for such tests and improvements.

iv. Please explain whether and how you have tested whether the "proactive automated detection systems" used have any negative impact for the exercise of the fundamental right to freedom of expression and information, in particular of marginalised groups. Please indicate any performance metrics and controls you have in place in this regard.

5. In its statement on the continued actions to protect the TikTok community during the Israel-Hamas conflict, TikTok states that it will continue to enforce its policies against violence and hate. However, several reports ([10]) inform that, allegedly, violent content

([10]) See, for instance, Attacks in Israel: Tech Against Terrorism monitors related violent tremist and terrorist content online ; In Israel-Hamas war, social media isn't neutral - The Washington Post; Families urged to delete TikTok, social media ahead of possible Israeli hostage videos | CNN Business; TikTok feature lets users pay to promote biased videos in searches for Israel and Palestine | Fortune; Why TikTok videos on Israel-Hamas war have drawn billions of views - The Washington Post; Fearing Hamas Execution Videos, Schools Urge Families To Delete TikTok And Instagram (forbes.com); Social media is awash in misinformation about Israel-Gaza war, but Musk's X is the most egregious - ABC News (go.com); Hamas Attacks, Israel Bombs Gaza and Misinformation Surges Online - bellingcat; TikTok Videos of Israel-Palestine Conflict Show New Role for the Platform - Bloomberg; Inside Israel-Gaza war propaganda, fake videos and misinformation online - The Washington Post.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

depicting hostage taking and other graphic videos are widely circulating on TikTok's platform, in violation of its own Community Guidelines.

i. Please share with the Commission any metrics and assessment of TikTok's enforcement measures with regard to the dissemination of illegal content as of 7 October 2023.

ii. Please explain whether TikTok put in place any additional measures as of 7 October 2023 to mitigate the risks related to wide dissemination of violent behaviours and hate speech in the context of the Hamas-Israeli conflict, and explain what these concrete measures are.

iii. Please provide evidence and a timeline of how TikTok has processed notices submitted by NGOs or national authorities and law enforcement with regard to content that is allegedly illegal or contrary to TikTok's terms and conditions and is related to the Hamas-Israeli conflict. Please refer to notices received since 28 August 2023 and explain the follow up given by TikTok to those notices and the time to process the notices before following up.

6. In its statement on the continued actions to protect the TikTok community during the Israel-Hamas conflict, TikTok states that it will continue to enforce its policies against harmful misinformation. In TikTok's response of 13 October to Commissioner's Breton's letter of 12 October 2023, TikTok states that "*We have already taken action in relation to fake/manipulated content and removed that which repurposes content from other conflicts or events (e.g. 'festivals and unrelated mass gatherings') and falsely claims that it represents current events. We do this by referring suspected content to our fact-checking partners and taking steps to prevent any removed content from resurfacing on the platform*". Please provide information and underlying documents related to the following questions:

i. How does TikTok identify which content is allegedly false/manipulated and need to be fact-checked? Please share with the Commission any metric and assessment of the effectiveness of the methods used to identify this type of content.

ii. Please provide the specific instructions given to TikTok's fact-checking partners in the context of the Hamas- Israeli conflict and the internal documents preparing these instructions.

iii. Please explain what are the measures TikTok has in place to prevent previously removed content from reappearing on the platform.

7. In its risk assessment report, TikTok states that one of the measures put in place to mitigate the risks, notably, for minors, stemming from illegal hate speech and for terrorist content, is the manual reviews of video content when it reaches certain levels of popularity in terms of number of video views, reducing the risks of violative content being shown in the "For You Feed" or otherwise widely disseminated.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_006401

i. Please describe the metrics used to determine the level of popularity of a video which triggers manual review.

ii. Please describe the number of the resources and qualifications of staff involved in this manual reviewing, including the language coverage and subject matter expertise of that staff.

8. In its risk assessment report, TikTok states that "*TikTok has taken a number of actions [...] to ensure it is able to respond quickly and effectively*" to terrorist and violent acts being livestreamed on its service. In its statement on the continued action to protect the TikTok community during the Israel-Hamas conflict, TikTok states that it is making "*temporary adjustments to policies*" to proactively prevent TikTok from being used for hateful or violent behaviour in the region, such as "*adding additional restrictions on LIVE eligibility as a temporary measure given the heightened safety risk in the context of the current hostage situation*".

i. Please explain concretely how TikTok has trained its moderation team to respond to content of this nature (e.g., what is the percentage of the moderators trained, what is the content of the training, etc.)

ii. Please provide the specific instructions given to TikTok's moderation team in the context of the Hamas-Israeli conflict and the internal documents preparing these instructions.

iii. In its risk assessment report, TikTok indicates that it "*augmented its internal risk detection and monitoring capabilities with supplementary information from partnerships with external threat detection partners*". Please indicate who are these "*external detection partners*" and the exact mission TikTok delegated to them, including their specific instructions.

iv. Please explain how these "*external detection partners*" are currently involved in the monitoring of the risks stemming from the livestreaming of violent acts in the context of the Hamas-Israeli conflict.

v. Please describe concretely, including through the provision of internal documents, the processes put in place to respond to any livestream of a violent act, such as execution of hostages and the "*additional restrictions on LIVE eligibility*" added in the context of the Hamas-Israeli conflict.

vi. Please indicate for how long the "*temporary adjustments*" will be made (if already determined).

vii. TikTok's risk assessment report presents the "*blocking words commonly used to search for violative content*" as a mitigation measure. Please indicate whether TikTok has already blocked such words in the context of the Hamas-Israeli conflict and the frequency of blocking. If yes, please list the blocked words (break down by languages).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_006402

a. Please provide any measurement or assessment of the effectiveness of these additional restrictions, the performance metrics and controls TikTok has put in place to prevent the livestreaming of violent acts such as execution of hostages.

b. Please explain whether and how you have tested whether this measure has any negative impact for the exercise of the fundamental right to freedom of expression and information, in particular of marginalised groups. Please indicate any performance metrics and controls you have in place in this regard.

9. In TikTok's response of 13 October to Commissioner's Breton's letter of 12 October 2023, TikTok states that it "*prohibit[s] synthetic media that contains material that has been edited, spliced, or combined (such as video and audio) in a way that may mislead a person about real-world events. We prohibit synthetic media that contains the likeness of any real private figure, anyone under 18, or which violates any other policy, including those on impersonation and misinformation. We will continue to remain vigilant in this area*".

i. Please describe how TikTok identifies "*synthetic media that contains material that has been edited, spliced, or combined (such as video and audio) in a way that may mislead a person about real-world events*".

ii. Please describe how TikTok identifies "*synthetic media that contains the likeness of any real private figure, anyone under 18, or which violates any other policy, including those on impersonation and misinformation. We will continue to remain vigilant in this area*".

iii. Please share with the Commission any metric, tests, testing results and assessment of the method(s) used for such identifications.

10. In TikTok's response of 13 October to Commissioner's Breton's letter of 12 October 2023, TikTok states that "*our policies prohibit covert influence operations that attempt to sway public opinion while misleading our community or systems. We have a dedicated team that works to monitor for, and remove, this content*".

i. Please describe how TikTok identifies "*covert influence operations that attempt to sway public opinion while misleading our community or systems".*

ii. Please provide information on the size (in FTE equivalent) of the dedicated team as well as information regarding the qualification of the staff involved in monitoring the presence of covert influence operation, including their language and subject-matter expertise.

11. Some media have reported that at least one prominent account that covers news from a Palestinian perspective was informed on Monday 9 October 2023 that it had been permanently banned on TikTok, before being reinstated following criticism that this

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_006403

ban amount to censorship of Palestinian voices ([11]). Please describe the internal metrics used to assess whether the content moderation decisions in the context of Hamas-Israel conflict have any negative impact on the fundamental right to freedom of expression and information, in particular of marginalised groups.

B. Questions related to the design of TikTok's recommender systems.

12. According to publicly available research, in the context of the ongoing conflict between Hamas and Israel, the TikTok's "You may like" section recommends harmful or disturbing content to its users, especially younger users, without any safeguards ([12]).

    i. Please explain what tests and studies TikTok has conducted to assess whether the recommender system(s) used for the "You may like" section might contribute to any of the systemic risks listed in Article 34(1), in particular the systemic risks related to public security, civic discourse, the protection of minors and/or the mental well-being of its users. In particular, please indicate whether TikTok has evaluated whether its recommender systems might expose users to violent and/or harmful content when they did not request it or express an interest in accessing this type of content providing also the specific features and outcomes of these tests. Please specify any performance metrics and controls you apply in this respect and explain which types of content and which recommender systems these measures apply to.

    ii. Please explain whether TikTok has operated any changes to the recommender system(s) and the measures taken to prevent the recommendation, recirculation, or amplification of illegal content related to the Hamas-Israeli conflict (e.g. illegal hate speech, terrorist content), as well as content that violates TikTok's policies applicable to this situation. Please refer to any changes and measures taken as of 28 August 2023 and specify which measures were applied on or after 7 October 2023 with respect to content available in the European Union. Please specify any performance metrics and controls TikTok applies in this respect and explain which types of content and which recommender systems these measures apply to.

    iii. Please explain whether and how you have tested the potential impact of TikTok's recommender system(s) on the fundamental right to freedom of expression and information. Please indicate any performance metrics and controls you have in place in this regard.

([11]) In Israel-Hamas war, social media isn't neutral - The Washington Post; TikTok, Instagram Target Outlet Covering Israel–Palestine Amid Siege on Gaza (theintercept.com)

([12]) How TikTok drags users into the digital war zone | auditing TikTok (tiktok-audit.com)

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

II. As regards TikTok's risk assessment and risk mitigation measures in relation to civic discourse and electoral process – compliance with Articles 34 and 35(1) of Regulation (EU) 2022/2065

A. Questions related to risk assessments

13. Which measures did TikTok take to reinforce its internal processes, resources, testing, documentation, or supervision of any of your activities in particular as regards detection of the systemic risk related to civic discourse and electoral processes?

    i. How did you take into account specific regional or linguistic aspects, for example as regards the use of your services in the context of the Slovak, Luxembourgish and Polish elections?

    ii. Please include an overview of the organisational set-up and human resources you have allocated for the Autumn 2023 general elections in Slovakia, Luxembourg and Poland, respectively, for assessing risks and mitigating them, expressed in terms of full-time equivalent (FTE) working within the company on election integrity for these elections specifically. This should include:

        a) The qualifications of these employees;

        b) Please explain how you have determined the number of FTEs and the qualifications of staff needed for each of the three elections taken as examples here.

    iii. Please provide details on the detection and actions taken against coordinated inauthentic campaigns in the context of each of the two elections, including any internal documents regarding your assessments.

14. On the basis of Article 34(3) of Regulation (EU) 2022/2065, the Commission requests TikTok to submit the documents underpinning the risk assessments of TikTok with respect to civic discourse and elections, notably all studies and evidence used for determining inherent risks, results of all tests, studies and internal assessments for the choice of the controls and assessing and monitoring their effectiveness, including documents on the testing of recommender systems as well as any further studies and internal documents underpinning the assessment of residual risks. Please submit the documentation used for the purpose of the risk assessment completed by 28 August 2023 pursuant to your obligation under Article 34(1) of Regulation (EU) 2022/2065.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_006405

B. Questions related to TikTok's policies as risk mitigation measures in relation to civic discourse and electoral processes systemic risks (Article 35(1) of Regulation (EU) 2022/2065) - please reply to the following questions focusing on the example of the recent Slovak, Luxembourgish and Polish elections, respectively, within the geographical scope of the European Union.

a. Incidents of election misinformation / voter suppression or dissuasion to vote.

15. Please provide the performance metrics and describe the internal controls you use for your efforts to detect infringements of your policy for 'Election Misinformation ([13]).

16. Please provide a timeline of the total daily content moderation decisions for your policies for 'Election Misinformation' as of 28 August 2023 or, if available, for the 6 months leading up to the respective elections.

17. Please report on the data and internal metrics you employ to measure the effectiveness of the policy for 'Election Misinformation' as a risk mitigation measure and provide any internal documents regarding the implementation of your policy.

18. Please explain the average time and notable outliers for processing notifications received under your terms and conditions with respect to violations of the 'Election Misinformation' policies.

b. Incidents of hate speech

19. Please provide the performance metrics and describe the internal controls you use for your efforts to detect infringements of your policy for hate speech.

20. Please provide a timeline of the total daily content moderation decisions for your policies for hate speech as of 28 August 2023 or, if available, for the 6 months leading up to the respective elections.

21. Please report on the data and internal metrics you employ to measure the effectiveness of the policy for hate speech as a risk mitigation measure and provide any internal documents regarding the implementation of your policy.

c. Synthetic and Manipulated Media

22. In its risk assessment report, TikTok refers to its policy on Synthetic and Manipulated Media as a risk mitigating measure. Please provide the performance metrics and describe the internal controls TikTok uses for its efforts to detect infringements of your policy for 'Synthetic and Manipulated Media'.

([13]) Community Guidelines | TikTok accessed on 10/10/2023

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_006406

23. Please provide a timeline of the total daily content moderation decisions for your policy for 'Synthetic and Manipulated Media' as of 28 August 2023 or, if available, for the 6 months leading up to the respective elections.

24. Please report on the overall effectiveness of the policy for 'Synthetic and Manipulated Media' as a risk mitigation measure and provide any internal documents regarding the implementation of your policy.

d. Other questions in relation to elections

25. In its risk assessment report, TikTok refers to the Fact-Checking Programme, with IFCN-certified independent fact-checking partners across Europe who assist TikTok in verifying, labelling or removing content, as a specific mitigation measure.

i. Please provide the performance metrics and describe the internal controls TikTok uses for the Fact-Checking Programme.

ii. Please provide a timeline of the total daily amount of content that received a Fact-Check label or was removed under this policy as of 28 August 2023 or, if available, for the 6 months leading up to the respective elections.

iii. Please provide any internal documents regarding the implementation of the Fact-Checking Programme during the respective elections.

26. In its risk assessment report, TikTok refers to in-app interventions for users to direct them to authoritative content, like the "elections hub".

i. Please provide the performance metrics and describe the internal controls TikTok uses for this "election hub".

ii. How many users interacted with the "election hub" and how much time was spent on the hub on average?

27. In its risk assessment report, TikTok mentions that it may conduct on and off app media literacy campaigns.

i. How big was the reach of this campaign for the Slovak, Luxembourgish and Polish elections?

ii. What are your performance metrics for this media literacy campaign besides the number of users reached?

28. Please provide an assessment of the lessons learned from the recent elections in Slovakia, Luxembourg and Poland, and please explain how you have incorporated these lessons into your risk mitigation measures for future elections. Please provide any internal documents regarding any evaluation of your risk mitigation strategy for the Slovak, Luxembourgish and Polish elections.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_006407

III. As regards the risk management framework in relation to specific topics (compliance with Articles 34 and 35 of Regulation (EU) 2022/2065)

A. Number of minors using TikTok

29. On the basis of Article 34(3) of Regulation (EU) 2022/2065, the Commission requests TikTok to submit the documents underpinning the risk assessment of TikTok with respect to the rights of the child in the form of statistics on the number of minors that use TikTok in the EU by age group (13-15, 16-17), by EU Member State, and in the EU as a whole, and as a proportion of the overall number of recipients of TikTok in each Member State and in the EU as a whole.

30. Please provide information on the method(s) used in the risk assessment to determine the number of users who are minors as well as their age group.

31. Please provide information, if possible concrete numbers, on the use of the service by minors below the age of 13, although prohibited by TikTok's terms and conditions.

B. Addictive behaviour

32. In its risk assessment report, TikTok states that it is employing mitigation measures to limit young users being exposed to "*repetitive content*" by diversifying it. However, the report does not explain concretely what these mitigation measures are, nor does it provide any evidence to support their effectiveness and proportionality. Furthermore, TikTok's Community Guidelines state that certain types of harmful content may be eligible for the For You Feed but will be interrupted to ensure it is not viewed "*too often*."

    i. Please describe concretely the measure(s) put in place to mitigate the systemic risk in relation to the "*repetitive content*" that you identified.

    ii. Please describe how TikTok determines what is "*too often*" in the context of its repetitive pattern policy.

    iii. Please provide any evidence including in the form of internal documents regarding tests regarding TikTok's efforts to detect repetitive patterns, their outcome and the conclusions drawn from the tests, including the timeline for such tests and improvements.

    iv. Please provide the performance metrics of the effectiveness of TikTok's mitigation measure(s).

33. The risk assessment report outlines different measures intended to limit the time young users spent on TikTok (Daily Screen Time Management dashboard and other screen time tools) as mitigation measures. However, the report does not explicitly identify (nor assess) as a potential systemic risk that minors might spend long periods of time on TikTok. Under article 34(3), the Commission requests the supporting documents of the

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_006408

risk assessment in the form of statistics on the amount of time that minors spend on TikTok (per age, per member state, the time of day).

34. TikTok's Community Guidelines, to which the risk assessment report referred, state that by default the daily screen time for minors between 13 and 17 years old is set to one hour and that users will "*be notified if [they] reach [their] daily screen time and [they] can choose to close the app or enter a preset passcode to return to TikTok.*" ([14]). Please provide the performance metrics of the effectiveness and proportionality of this measure.

35. Please provide any evidence including in the form of internal documents regarding tests regarding the development of TikTok's efforts to Daily Screen Time Management dashboard and other screen time tools as mitigation measures, their outcome and the conclusions drawn from the tests, including the timeline for such tests and improvements.

C. In relation to the spread of dangerous challenges on TikTok

36. In its risk assessment report, TikTok acknowledges that the dissemination of "*Online Challenges*" on its platform might constitute a risk for its users. It also indicates that it updated its Community Guidelines and made changes to its moderation practices to protect users against such risk. TikTok's Community Guidelines, indicate that showing or promoting dangerous challenges is not allowed ([15]). However, media reports ([16]) reveal that such dangerous challenges (the recent hot chip challenge, the blackout challenge, bridge jumping, Kia boys, jeu de virgule, "French scar" challenge) that can involve physical or psychological damage to individuals or property appear on TikTok.

    i. Please explain in detail the method(s) used to enforce TikTok's policies on dangerous challenges.

([14]) Screen time | TikTok Help Center [accessed 11th October 2023]

([15]) Community Guidelines | TikTok [accessed 11th October 2023]

([16]) See, for example, Hot-Chip-Challenge – Infos und Tipps für Eltern und Lehrkräfte - klicksafe.de; Family says Worcester teen died after spicy chip challenge (boston.com); Disturbing TikTok 'Bone Smashing Challenge' Encourages Risky Self-Harm for Beauty (viewusglobal.com); 'Bone Smashing' TikTok Trend, Here Are Dangers Of Hammering Your Face (forbes.com); Who are the "Kia Boyz"? How TikTok fueled an epidemic of car thefts | Ars Technica; TikTok 'Kia Challenge' Fuels Rise in Using USB Cables to Steal Cars (insider.com); TikTok 'French Scar' challenge triggers safety probe in Italy | TechCrunch; Italy's Antitrust probes TikTok over 'French scar' dangerous content | Reuters; Ouders en experten waarschuwen voor gevaarlijke challenge op Tiktok: "Je kan in coma raken of erger" | Het Nieuwsblad; Le 'jeu de la virgule', défi dangereux, se répand dans les cours d'école | TV5MONDE - Informations; "Jeu de la virgule" : "blessée par l'un de ses camarades", une nouvelle écolière hospitalisée - lindependant.fr; Le 'jeu de la virgule', défi dangereux, se répand dans les cours d'école - rtbf.be; Le "jeu de la virgule", ce nouveau jeu dangereux qui sévit dans les cours d'école: un élève de CM1 blessé aux cervicales à Besançon (francetvinfo.fr); 10 Bad TikTok Challenges That Went Too Far (makeuseof.com); My son died, 15, in a social media challenge gone horrifically wrong - TikTok 'jokes' have to stop, they're killing kids | The Sun

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

ii. Describe the method(s) of identifying and classifying dangerous challenges.

iii. Please provide any evidence including in the form of internal documents regarding tests regarding the identification and classification of dangerous challenges and the measures to prevent showing or promoting them, their outcome and the conclusions drawn from the tests, including the timeline for such tests and improvements.

iv. According to recent media reports ([17]), a new challenge, referred to in French as "jeu de la virgule", has been spreading as a challenge on TikTok. Please describe how TikTok is responding to this, considering that, according to those reports, the challenge has resulted in severe body harm in the form of cervical damage.

v. Please share with the Commission any metric and assessment of the effectiveness of TikTok's Community Guidelines and moderation practices to address the risks posed by the dissemination of dangerous challenges on its service.

37. The report conducted by Praesidio Safeguarding in 2021, annexed to TikTok's risk assessment report ([18]), explores challenges related to self-harm and suicide, as well as hoax challenges, but does not cover challenges that result in theft or damage of property. TikTok's policy on dangerous challenges ([19]) does not cover those that result in theft or damage of property. Please explain the reasons for which, in designing this policy around dangerous challenges, this was not included.

38. In its risk assessment report, TikTok presents the fact that content deemed to be a dangerous challenge is not eligible to be featured on the For You Feed (FYF) as a mitigation measure to the risk of dangerous challenges. Please provide performance metrics on the effectiveness of this mitigation measure.

Thank you for your cooperation.

([17]) See, among others, https://france3-regions.francetvinfo.fr/bourgogne-franche-comte/doubs/besancon/jeu-de-la-virgule-un-jeu-dangereux-un-eleve-de-cm1-blesse-aux-cervicales-a-besancon-2850086.html; Le «jeu de la virgule»: un défi dangereux qui peut entraîner une paralysie, préviennent les pédiatres - La Voix du Nord; Le 'jeu de la virgule', défi dangereux, se répand dans les cours d'école - rtbf.be

([18]) Risk assessment report pgs. 26-27 and Exploring-effective-prevention-education-responses-to-dangerous-online-challenges-English-UK-compressed-1.pdf (praesidiosafeguarding.co.uk)

([19]) https://www.tiktok.com/community-guidelines/en/mental-behavioral-health/#3

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_006410

ANNEX II

INSTRUCTIONS FOR REPLYING

i. The reply and relevant documents submitted under Regulation (EU) 2022/2065 shall be submitted in a format allowing the Commission services to process them electronically and, in particular, enabling their digitisation and character recognition. The reply must clearly indicate the number of the file and the date. The reply should be page-numbered, and the paragraph numbered.

ii. Please submit your reply by sending it to the functional mailbox of the Directorate General for Communications Networks, Content and Technology at: CNECT-DSA-Registry@ec.europa.eu.

iii. The electronic version of your reply shall be considered by the Commission services as an original reply. There is therefore no need to submit a hard-copy in addition to the electronic version of the reply.

Electronically signed on 19/10/2023 11:00 (UTC+02) in accordance with Article 11 of Commission Decision (EU) 2021/2121

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_006411

Exhibit 291

EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR COMMUNICATIONS NETWORKS, CONTENT AND TECHNOLOGY

Platforms Policy and Enforcement
Digital Services

Brussels, 14 March 2024

SENSITIVE (*)

Redacted via email:
Redacted tiktok.com

Subject: Request for information pursuant to Article 67(1) and Article 34(3) of Regulation (EU) 2022/2065 regarding TikTok's compliance with Articles 34 and 35 of Regulation (EU) 2022/2065 in relation to TikTok

Dear Redacted,

On 25 April 2023, TikTok was designated as a very large online platform ('VLOP') pursuant to Article 33(4) of Regulation (EU) 2022/2065 of the European Parliament and the Council of 19 October 2022 on a Single Market for Digital Services. On the following day, that designation decision was notified to TikTok Technology Limited, and an acknowledgement of receipt was received by the Commission on 27 April 2023. A copy of the Designation Decision was sent to ByteDance Limited, the company controlling the group of legal entities to which TikTok Technology Limited belongs. Consequently, pursuant to Articles 33(6) and 92 of Regulation (EU) 2022/2065, that regulation started to apply to TikTok as from 28 August 2023.

Pursuant to Article 67 of Regulation (EU) 2022/2065 the Commission may, by simple request, require a provider of VLOP to provide information on a suspected infringement of that Regulation. In accordance with Article 65(1), the Commission services may make such a request before initiating formal proceedings.

For the reasons set out below, the Commission suspects that TikTok Technology Limited and ByteDance Limited ('TikTok') may not be compliant with certain obligations under Regulation (EU) 2022/2065. By the present letter, the Commission therefore requests certain information from you pursuant to Article 67(1) of that Regulation, as well as the documentation supporting TikTok's risk assessment pursuant to Article 34(3) of that Regulation.

Pursuant to Article 34(1) of Regulation (EU) 2022/2065, providers of VLOPs are obliged to diligently identify, analyse, and assess any systemic risks in the Union stemming from the design or functioning of their service and its related systems, including "algorithmic systems", or from the use made of their services. Pursuant to Article 34(2) of Regulation (EU) 2022/2065, in their risk assessment providers of VLOPs must take into account some

(*) Distribution only on a need-to-know basis. Do not read or carry openly in public places. Must be encrypted in transmission and stored securely, where possible by using encrypted storage. Destroy copies by shredding or secure deletion. Full handling instructions: https://europa.eu/!db43PX

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

specific factors that can influence any of the systemic risks referred to in Article 34(1), including the design of any relevant algorithmic system, their content moderation systems, the applicable terms and conditions, and the intentional manipulation of their service.

Pursuant to Article 35(1) of Regulation (EU) 2022/2065, providers of VLOPs are obliged to put in place reasonable, proportionate, and effective mitigation measures tailored to the specific systemic risks identified. Such measures may include: (i) adapting their terms and conditions, (ii) testing and adapting their algorithmic systems; (iii) reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risks; (iv) taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information; (v) ensuring the possibility to clearly distinguish generated or manipulated images, audio or videos.

On 29 September, TikTok submitted its risk assessment report to the Commission. The report also includes information about mitigation measures under Article 35(1) of Regulation (EU) 2022/2065.

The Commission has reasons to suspect that TikTok may not be compliant with Articles 34 and 35 of Regulation (EU) 2022/2065 with particular reference to specific risks related to the dissemination, within its service, of content that was produced by generative artificial intelligence (i.e., Generative AI content), including text, images, videos, audio or other synthetic content. As explained in the questions in Annex I to this letter regarding compliance with those provisions, the Commission suspects, on the basis of the information provided in TikTok's risk assessment report, that TikTok has not diligently identified, analysed and assessed all the systemic risks stemming from Generative AI content, in line with Article 34 of Regulation 2022/2065, and that the mitigation measures outlined in that report are not reasonable, proportionate and effective to address those risks, as required by Article 35 of Regulation (EU) 2022/2065.

First, TikTok's risk assessment report does not provide sufficient information on the results related to the identification, analysis, and assessment of the specific risks stemming from Generative AI content, including its use, misuse, and dissemination on TikTok, in relation to the dissemination of illegal content, any actual or foreseeable negative effects on the exercise of fundamental rights, including human dignity, non-discrimination, protection of personal data, consumer protection and intellectual property, to gender-based violence and the protection of public health, public security, minors and mental and physical well-being, and on civic discourse and electoral processes. Indicatively, in the risk assessment report, under the section "Emerging risks" TikTok states that "*TikTok is closely monitoring the risks of generative AI and considered its impact in a number of areas contained in this Report. TikTok will continue work to improve its methods for detecting synthetic media at scale.*" ([1]). Further on, the risk assessment report lists some "*proactive measures to mitigate generative AI risks in a reasonable and proportionate way*" and states that "*The current state of the art of detecting synthetic media at scale is still limited. However, TikTok continues to engage with its peer platforms on current and upcoming trends to further inform its approach to monitoring this emerging risk.*". Second, the report lacks a

([1]) Page 8 of the report.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

description of the measures put in place to mitigate the identified systemic risks stemming from Generative AI content on TikTok, including an assessment of their reasonableness, proportionality and effectiveness.

Pursuant to Article 34(3) of Regulation (EU) 2022/2065, providers of VLOPs are required to preserve the supporting documents of the risk assessments for at least three years after the performance of the risk assessments and to communicate them to the Commission, upon request. That provision further provides that the Commission may request such supporting documentation on which the platform relied to draw up its risk assessment report.

The Commission adopted decision C(2024) 1158 final initiating proceedings on the basis of Articles 34(1) and (2), 35, 28(1), 39(1) and 40(12). The present request for information does not fall within the scope of decision.

Based on Articles 67(1) and 34(3) of Regulation (EU) 2022/2065, the Commission requests TikTok to respond to the questions set out in Annex I to this letter regarding compliance with Articles 34 and 35 of that regulation.

On these bases, the Commission requests TikTok to reply to the questionnaire attached and to submit the supporting documents requested in Annex I, in order to enable the Commission to further assess whether TikTok complies with Articles 34 and 35 of Regulation (EU) 2022/2065.

a) Concerning the requests under section I, please reply **no later than 5 April 2024;** Concerning the requests under section II, please reply **no later than 26 April 2024.**

Please submit your reply by sending it to the functional mailbox of the Directorate General for Communications Networks, Content and Technology at CNECT-DSA-Registry@ec.europa.eu.

Unless you specify otherwise, your response shall be treated as non-confidential. If you request that any piece of information supplied in your response should not be published in any public version of Commission documents related to TikTok's compliance with Regulation (EU) 2022/2065 or otherwise disclosed to third parties, please submit both a confidential and a non-confidential version of your reply and provide reasons for any confidentialities claimed.

Any personal data submitted in your reply will be processed in compliance with Regulation (EU) No 2018/1725 of the European Parliament and of the Council ([2]). Please see here the privacy statement on the application of the regulatory framework, including processing of notifications, monitoring of compliance, supervisory activities, and other activities under the Digital Services Act.

If you consider that some documents included in your reply (or parts of some documents) contain personal data revealing racial or ethnic origin, political opinions, religious or

([2]) Regulation (EU) 2018/1725 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC (OJ L 295, 21.11.2018, p. 39).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_018317

philosophical beliefs, or trade union membership or genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation ([3]) (i.e. special categories of personal data), please provide those documents in a separate submission identified as "SPECIAL CATEGORIES OF PERSONAL DATA", linking each document to the question number to which the document constitutes a reply. Please highlight the special categories of personal data in the document and provide an explanation as to the specific category in case this is not immediately obvious from the information highlighted.

Please note that, pursuant to Article 74(2) point (a) of Regulation (EU) 2022/2065, the Commission may impose on a provider of a VLOP fines not exceeding 1 % of the total annual income or worldwide turnover in the preceding financial year where they intentionally or negligently supply incorrect, incomplete or misleading information in response to a simple request for information pursuant to Article 67.

If you have any further questions, please address them by e-mail to CNECT-DSA-Registry@ec.europa.eu, clearly indicating the subject of this letter and the reference number.

Yours faithfully,

Redacted

Enclosures:

- Annex I: Questionnaire
- Annex II: Instructions for replying

c.c.: Redacted tiktok.com; Redacted bytedance.com; Redacted tiktok.com; Redacted tiktok.com

([3]) See Article 10(1) of Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data (EU-DPR), and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC (OJ L 295, 21.11.2018, p.39), which is identical to Article 9(1) of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (GDPR) (OJ L119, 4.5.2016, p.1).

ANNEX I

QUESTIONNAIRE

TikTok

DEADLINE FOR REPLY:

For Section I: 5 April 2024

For Section II: 26 April 2024

1. BUSINESS SECRETS AND OTHER CONFIDENTIAL INFORMATION

Does your reply contain any parts that you regard as business secrets or other confidential information that should not be published or otherwise disclosed to third parties?

YES ☐ NO ☐

If yes, please submit both a confidential and a non-confidential version of your reply and provide reasons for any confidentialities claimed.

2. QUESTIONS

Format of the response

See Annex II. Please provide your response in Word or Excel.

Specific questions

In several questions you are requested to provide information on performance metrics and internal controls:

- *By performance metrics, the Commission means in particular, but not only, quantitative performance indicators that help determine objectively whether a given policy, action or measure is successful or not, or that allow the determination of progress towards reaching the objectives of the measure.*

- *By internal controls, the Commission means any type of measure, including business process and tests conducted and organisation arrangement, that aims to monitor and ensure the effectiveness and correct functioning of the measures put in place.*

Where your organisation is using different terminology, please explain the relevant organisational set-ups that are closest to the requested concepts.

Period of time covered

Unless otherwise specified, the period of time covered by this request for information is the period of time starting as from the date from which Regulation 2022/2065 started to apply to TikTok, i.e. as from 28 August 2023, until the closest possible date for which data/information were available before the submission of your reply. This period is

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_018319

hereinafter referred to as 'the relevant period'. Where relevant, please do not limit your reply to the situation at the time of submitting your response but provide details on any changes which may have occurred since 28 August 2023 or the date otherwise specified in the specific question.

Where you have already provided the requested information based on the previous requests sent by the Commission, please provide updated information starting from the date your previous reply was limited to. Please indicate clearly what has changed since the previous reply.

As regards compliance with Article 11(1)

Please state which is your designated single point of contact to enable communications directly, by electronic means, with Member States' authorities, the Commission and the Board.

I. As regards the risk assessment and risk mitigation measures (compliance with Articles 34 and 35 of Regulation 2022/2065) in regard to risks stemming from Generative AI in relation to civic discourse and electoral processes

Very large online platforms (VLOPs) have become important venues for electoral processes. Given the number of elections in 2024 ([4]), this issue takes on particular importance and urgency. In this context, Generative AI technology presents new risks, as evidenced by its dedicated section (3.3) in the recent guidelines on the integrity of electoral processes, on which the Commission sought views ([5]).

In light of this, the documentation at the Commission's disposal raises concerns that TikTok has not diligently assessed and mitigated potential systemic risks stemming from Generative AI, particularly in relation to any actual or foreseeable negative effects on civic discourse and electoral processes.

On this basis, the Commission requests the following information. Please give due consideration that this request has an urgent character, and that, except for duly justified circumstances, your company will not be granted an extension to the original deadline.

For each provided answer, pursuant to Article 34(3) of Regulation (EU) 2022/2065, submit all supporting documents. Include measurements of the performance and effectiveness of all relevant measures, including internal documents, presentations to management, internal exchanges/emails about effectiveness of the measures, meeting notes, and other measures considered and discarded.

([4]) Indicatively, parliamentary elections in March in Portugal, in June in Belgium, in July in Croatia, and September in Austria, in addition to the European Parliament elections in June.

([5]) Commission is gathering views on draft DSA guidelines for election integrity | Shaping Europe's digital future (europa.eu)

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_018320

A. In relation to the dissemination of Generative AI content related to risks for civic discourse and electoral processes

1. Please set out in detail internal guidelines, policies, and practices related to Generative AI content in TikTok. This includes content generated on other platforms or within TikTok's services. Additionally, please describe in detail how TikTok enforces such policies.

2. Please describe any partnerships with other providers or other consortia that TikTok is a member of, related to Generative AI. Please describe the actions (e.g., meetings, technical exchanges, shared developments), outcomes (e.g., new technologies / systems, new processes), and underlying documents (e.g. meeting notes, emails/exchanges) of these partnerships and consortia and describe how they have been implemented within TikTok.

3. Please describe in detail the risk assessment concerning and any mitigation measure taken to detect Generative AI content on the service, especially with regard to synthetic or manipulated images, audio or videos that appreciably resemble existing persons, objects, places, entities, events, or depict events as real that did not happen or misrepresent them (i.e., deepfakes). In doing so, please devote special attention to content concerning politicians, candidates, political parties, false depiction of events, election polls, information on how, where and when to vote, contexts or narratives, or any other relevant situation for the electoral political context.

4. TikTok's risk assessment report ([6]) states that "*TikTok does allow synthetic media of public figures as long as the content is not used for an endorsement or is otherwise violative*". Please describe in detail how TikTok assessed the risk of this policy, and how TikTok defines and detects "*public figures*", "*endorsement*", and "*otherwise violative*". Please describe in detail systems, and provide internal documentation of performance metrics, experiments and tests, qualitative and quantitative results (including on accuracy, precision or other relevant metrics), internal processes, internal controls, and other input to detect these.

5. Please describe in detail the risk assessment concerning and the content moderation actions taken on Generative AI content, detected within the service, including removals, demotions, etc. by Member State or EU language and at the EU level. In doing so, please devote special attention to content concerning politicians, candidates, political parties, false depiction of events, election polls, information on how, where and when to vote, contexts or narratives, or any other relevant situation for the electoral political context.

6. Please describe in detail the actions taken on Generative AI content, detected within the service, including labelling, watermarking, etc., additionally clarifying whether TikTok provides users with interfaces and tools to add labels to Generative AI content and whether such labels are retained once the content is shared by other users on the platform.

([6]) Page 35.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

7. Please describe in detail any crisis response protocols, policies, or procedures put in place to address risks that may arise from the fast and viral dissemination of Generative AI content (e.g., non-watermarked deepfake content virally spreading through the service in the EU within a short timeframe). Please describe the processes in place to detect and address such high-impact risks, including the criteria used to activate responses, the response mechanisms, de-escalation procedures, the organisational set-up of response teams, and specific events where these processes were triggered in during the reporting period, along with the respective responses.

II. As regards the risk assessment and risk mitigation measures (compliance with Articles 34 and 35 of Regulation 2022/2065) in regard to risks stemming from Generative AI (other than those already addressed in the previous section devoted to civic discourse and electoral processes)

The Commission has reasons to suspect that TikTok may not be compliant with Articles 34 and 35 of Regulation (EU) 2022/2065 with particular reference to specific risks stemming from Generative AI content (e.g., deepfakes), in relation to the dissemination of illegal content, on any actual or foreseeable negative effects on the exercise of fundamental rights, including human dignity, non-discrimination, protection of personal data protection, consumer protection and intellectual property, gender-based violence and the protection of public health, public security, minors and mental and physical well-being.

For this reason, the Commission considers that additional information and clarifications are required to assess if the above-mentioned risks have been diligently assessed and if reasonable, proportionate and effective mitigation measures are put in place, in line with the relevant obligations under Articles 34 and 35 of Regulation (EU) 2022/2065 respectively. On this basis, the Commission requests the below information.

For each provided answer, pursuant to Article 34(3) of Regulation (EU) 2022/2065, submit all supporting documents. Include measurements of the performance and effectiveness of all relevant measures, including internal documents, presentations to management, internal exchanges/emails about effectiveness of the measures, meeting notes, and other measures considered and discarded.

A. General questions in relation to the dissemination and creation of Generative AI content on TikTok

1. Please set out in detail TikTok's internal guidelines, policies, and practices related to Generative AI content. This includes content generated on other platforms or within TikTok's services. Additionally, please describe in detail how TikTok enforces such policies.

2. Please describe any partnerships with other providers or other consortia that TikTok is a member of, related to Generative AI. Please describe the actions (e.g., meetings, technical exchanges, shared developments), outcomes (e.g.,

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

new technologies / systems, new processes), and underlying documents (e.g. meeting notes, emails/exchanges) of these partnerships and consortia and describe how they have been implemented within TikTok.

B. In relation to the dissemination of Generative AI content on TikTok

1. Please describe in detail systems, and provide internal documentation of performance metrics, experiments and tests (including A/B testing), qualitative and quantitative results (including on accuracy, precision or other relevant metrics), internal processes, internal controls, and other input (including from recipients of the service, independent experts, civil society organisations, and scientific insights, as referred to in recital 90 of Regulation (EU) 2022/2065) to detect Generative AI content on the service, especially with regard to synthetic or manipulated images, audio or videos that appreciably resemble existing persons, objects, places, entities, events, or depict events as real that did not happen or misrepresent them (i.e., deepfakes).

2. Please describe in detail the actions taken by TikTok on Generative AI content, detected within the service, including content moderation actions (e.g., removals, labelling, watermarking, etc.). Please provide a breakdown of the number of Generative AI items of content that was moderated across TikTok's services for the period covered per policy violation, per type of content (e.g., images, text, audio, and other synthetic media), and reason. Provide evidence about studies, experiments and tests performed by TikTok to assess the effectiveness of these actions, especially regarding deepfakes.

3. Please clarify whether TikTok provides users with interfaces and tools to add labels to Generative AI content. Please provide summaries, including quantitative data, of experiments and tests done to measure the effectiveness, and ease-of-use of these interfaces and tools. Please clarify whether such labels are retained once the content is shared by other users on the platform. Please explain whether metadata related to Generative AI content on TikTok is changed in any way and the relevant reasons for such actions.

4. Please describe in detail policies or procedures in place to address issues that may arise related to the viral dissemination of Generative AI content (e.g., non-watermarked deepfake content virally spreading through the service in the EU within a short timeframe). Please describe processes in place to detect such issues, criteria used to activate responses, response mechanisms, de-escalation procedures, and specific events where these processes were triggered in during the period covered, along with the respective responses.

C. In relation to the creation of Generative AI content on TikTok

1. Please provide a detailed description of all Generative AI systems made available by TikTok in the European Union, including their design, functioning and the underlying model. Indicatively, the Commission takes note of the 'AI

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

greenscreen' ([7]), 'Bold Glamour filter' ([8]), and 'TikTok Creative Assistant' ([9]). Please also provide a description of Generative AI systems that TikTok is currently working on that will be made available in the European Union in the near future.

2. Please explain how TikTok has, or plans to, diligently identify, analyse and assess the risks stemming from the design, functioning, use and misuse of the aforementioned Generative AI systems taking into consideration their severity and probability, for each of the following systemic risks:

   a. the dissemination of illegal content;

   b. any actual or foreseeable negative effects for the exercise of fundamental rights, in particular the right to (i) human dignity, (ii) protection of personal data, (iii) non-discrimination , (iv) consumer protection, and (v) intellectual property;

   c. any actual or foreseeable negative effects in relation to gender-based violence, public security, the protection of public health and minors and serious negative consequences to the person's physical and mental well-being.

3. Unless already provided above, please provide all documents underpinning the risk assessments with respect to questions C.1 and C.2, pursuant to Article 34(3) of Regulation (EU) 2022/2065.

4. Where applicable in reference to each specific aforementioned Generative AI system, regarding the specific mitigation measures put in place pursuant to Article 35 of Regulation (EU) 2022/2065, please describe in detail systems, and provide internal documentation of performance metrics, experiments and tests, qualitative and quantitative results, internal processes, internal controls put in place by TikTok to:

   a. align the output of the aforementioned systems with the terms and conditions, policies, and codes of conduct of the service;

   b. ensure that the aforementioned systems' outputs do not appreciably resemble existing persons, objects, places, entities, events, or depict events as real that did not happen or misrepresent them (i.e., deepfakes).

   c. ensure that Generative AI content is clearly distinguishable for users (through e.g., watermarking);

   d. ensure that Generative AI systems' output relies on reliable sources to limit the effects of 'hallucinations' by Generative AI;

([7]) https://www.theverge.com/2022/8/15/23306101/tiktok-ai-greenscreen-filter-text-to-image-mainstream

([8]) https://www.scmp.com/yp/discover/advice/living/article/3221222/effects-tiktoks-bold-glamour-filter-how-teens-picture-themselves-their-body-image

([9]) https://www.tiktok.com/business/en/blog/tiktok-creative-assistant-smart-ai-tool

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

e. warn users about potential errors in content created by Generative AI systems;

f. monitor the performance of Generative AI systems by, for example, setting and tracking appropriate safety and accuracy metrics;

g. ensure an adequate level of safety of its Generative AI systems by integrating safeguards, such as prompt and output classifiers, system-wide prompts, content moderation and other filters to prevent prompts and outputs that go against TikTok's terms and conditions.

5. Please describe and provide the related documentation about any red-teaming exercises conducted and documented by TikTok with both internal teams and external experts before making the aforementioned Generative AI systems available to users.

6. Please clarify whether TikTok consulted with recipients of the service, independent experts, civil society organisations, scientific insights, etc., in assessing and mitigating the risks of the aforementioned Generative AI systems and features before making them available to users. If yes, please provide descriptions and summaries of these consultations and insights.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_018325

ANNEX II

INSTRUCTIONS FOR REPLYING

The reply and relevant documents submitted under Regulation (EU) 2022/2065 shall be submitted in a format allowing the Commission to process them electronically and enabling their digitisation and character recognition. The reply must clearly indicate the number of the file and the date. The reply should be page-numbered and the paragraphs numbered.

Please submit your reply by sending it to the functional mailbox of the Directorate General for Communications Networks, Content and Technology at: CNECT-DSA-Registry@ec.europa.eu.

The electronic version of your reply shall be considered by the Commission as an original reply. There is therefore no need to submit a hard-copy in addition to the electronic version of the reply.

Electronically signed on 14/03/2024 11:07 (UTC+01) in accordance with Article 11 of Commission Decision (EU) 2021/2121

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_018326

Exhibit 292

From: Jason Pielemeier [redacted]@globalnetworkinitiative.org]
Sent: 5/27/2024 1:21:06 AM
To: [redacted]@ec.europa.eu; [redacted]@unisg.ch; Lene Wendland [redacted]@un.org]; [redacted]@google.com]
CC: Isabel Laura Ebert [redacted]@un.org]; Hilary Ross [redacted]@globalnetworkinitiative.org]; Radka Sibille [EEAS EUROPA] [redacted]@eeas.europa.eu]; Rumman Chowdhury [redacted]@humane-intelligence.org]
Subject: Re: Preparing for our panel on "Algorithmic Risk Assessments, Audits, and Fundamental Rights"

Hi All,
I'm on my way to Geneva as I write and am very excited to see you all this week. To help move us along in our panel prep, I've drafted a run-of-show and some questions for each of you and pasted them below. Given time constraints, I think we will only be able to ask **one direct question** to each panelist, which I'd like to ask you to prepare to **respond to in less than 5 minutes**. After that, I will **ask a single question to the full panel and invite you each to speak to it for no more than 3 minutes**. Time permitting, we will also take a couple of questions from the audience. Please let me know if you have feedback or questions.

Best,
Jason

Run of Show

14.00: Opening remarks: Ambassador Lotte Knudsen, Head of EU Delegation to the UN in Geneva, Peggy Hicks, Director, Thematic Engagement, Special Procedures and Right to Development Division, United Nations Human Rights

14.15-15.10: Opening panel

- 14:15-20: Welcome and Introduction
- 14:20-14:40: Individual questions to panelists (5 minute responses):
    - Juha Heikkila is Adviser for Artificial Intelligence in the European Commission's Directorate-General for Communications Networks, Content and Technology.
        - Juha, you have been working within the EU context on tech policy broadly, and AI specifically for many years. It would be wonderful to hear your thoughts on why the EU has focused as much as it has on the concept of "risk and impact assessments" and how you hope that these exercises will help protect fundamental rights?
    - Lene Wendland is Chief of the Business and Human Rights Unit in UN Human Rights, where among other things she leads the B-Tech Project, which works to apply the UN Guiding Principles on Business and Human Rights (UNGPs) to digital technologies.
        - Lene you worked with Prof. John Ruggie on the development of the UNGPs, which helped to center the concept of "human rights due diligence" as a core expectation for business. Could you tell us about the logic behind that move and how you see the relationship between HRDD and other related concepts like "risk assessments" and "impact assessments"?
    - Mariana Valente is Associate Director at InternetLab, a Brazilian think tank focused on human rights and digital technologies, as well as a professor at the University of Saint Gallen.
        - Mariana, Brazil is one of several countries that has taken serious steps toward developing legislative frameworks to address the potential risks of both digital platforms and AI. It would be great to hear about where things stand currently in

Confidential - Not For Public Release

Brazil and the similarities and differences you see between the Brazilian approach and what we have been seeing in the EU?
    - [redacted] is Global Head of Human Rights at Google, and a GNI Board member
        - [redacted] Google has several services that have been designated as VLOP and VLOSE under the DSA, and is also subject to risk assessment and due diligence obligations under the UK's Online Safety Act, Basic Online Safety Expectations in Australia, India's Intermediary Guidelines, among other regulatory frameworks. Google is also a part of the B-Tech Community of Practice, as well as a founding member of GNI about to embark on its fifth independent GNI assessment. Could you tell us about how non-regulatory frameworks like the UNGPs and GNI Principles have helped Google understand, identify, and address human rights risks, and how you see the commitments, systems, and processes you've developed under those frameworks being relevant to or in tension with emerging regulatory approaches that share the same broad objectives?
- 14:40-14:55: Question to all panelists (3 minute responses):
    - We've heard about how risk assessment and due diligence are well established concepts in the business and human rights and tech policy spaces. Now I'd like to ask you each to speak briefly about how you think such concepts translate or don't translate to the specific challenges related to algorithmic systems and artificial intelligence.
    - Order: [redacted] Mariana, Lene, Juha
- !4:55-!5:10: Audience questions + one round of panelist responses (2 minutes each)

On Mon, May 20, 2024 at 10:27 PM Jason Pielemeier <[redacted]@globalnetworkinitiative.org> wrote:

Dear [redacted] Juha, Lene, and Mariana,

Thank you all very much for your willingness to be part of this fantastic panel next Thursday. I would be remiss if I didn't also tip my hat to Hilary, Isabel, Radka, and Rumman for all of their great work putting the event together. GNI is thrilled to be a part of this and I am honored to moderate the panel.

After opening remarks, we will have 55 minutes for the panel, which includes time for audience Q&A. I'm attaching the concept note and pasting below the eight questions that the event is framed to explore. We could use some combination of these to guide our discussion. Alternatively, if there are related questions that you would prefer to be asked, please feel free to suggest them. I don't think that we'll have time to organize a call together ahead of the event, but if any of you would like to connect one-on-one beforehand, I am happy to try to find time.

Otherwise, if you have any questions or concerns, please let us know.

Looking forward to seeing you all next week,
Jason

The landscape of AI regulation:

- What are key global trends with regard to regulation requiring tech companies to assess human rights risks?

Identifying suitable methodologies to assess AI products and services with regard to human rights risks:

- How can stakeholders (including engineers) encourage comparable AI risk assessment and auditing benchmarks?
- What might appropriate methodologies for AI auditing look like and what data is needed to perform accountable AI audits?

Confidential - Not For Public Release

Requirements for effective oversight/enforcement and involvement of experts:

- What is the role of enforcing/supervisory mechanisms? (e.g. EU AI Office)
- How can civil society and academia most meaningfully engage around these processes?
- How can AI risk assessments and audits be used by companies and external stakeholders to ensure accountability and catalyse change?

--

GLOBAL NETWORK INITIATIVE

Jason Pielemeier | Executive Director
tel: +[redacted] | tw: @theGNI

1325 G St. NW, Suite 500
Washington, D.C. 20005
GlobalNetworkInitiative.org

--

GLOBAL NETWORK INITIATIVE

Jason Pielemeier | Executive Director
tel: +[redacted] tw: @theGNI

1325 G St. NW, Suite 500
Washington, D.C. 20005
GlobalNetworkInitiative.org

Confidential - Not For Public Release

Exhibit 293

Event Concept Note

Algorithmic Risk Assessments, Audits, and Fundamental Rights

Thursday 30 May 2024, 14:00-16:30 CET
Venue: European Union Delegation, Rue du Grand-Pré 66, 1202 Genève

Overview

With a view to effectively addressing and preventing violations and abuses of human rights online, often facilitated by ever-increasingly powerful AI systems, numerous **tech regulations have been emerging to set up safeguards in the whole life-cycle of technologies. Companies are required to perform human rights due diligence and risk assessments**, along with related transparency and audit requirements relating to digital technologies, including AI. In particular, the EU's Digital Services Act (DSA) relies on risk assessment, mitigation, auditing, and data transparency practices to hold large digital services accountable, in a manner that protects fundamental rights. Also following a risk-based approach, the recently adopted EU AI Act, as the first-ever comprehensive legal framework on AI worldwide, sets rules to foster trustworthy AI, by ensuring that AI systems respect fundamental rights, safety, and ethical principles and by addressing risks of very powerful and impactful AI models. Similar efforts have intensified in other regions as well, including Latin America with various countries starting to prepare their own regulations on AI, and Africa with the ongoing work on AI by the African Union Commission.

Ideally, these new regulatory frameworks will be informed by decades of voluntary practices – transparency reporting, human rights risk assessment, and auditing – developed to encourage responsible business conduct in line with the UN Guiding Principles on Business and Human Rights (UNGPs). However, these regulatory developments necessitate the convergence of traditional auditing and assessment processes with technical audits. For oversight and enforcement, often companies are now required to share data and code, enabling auditors to evaluate algorithms and datasets. **This is a promising development on the path to enabling accountability and AI for Good, while safeguarding against adverse impacts on human rights.**

Yet, many questions and challenges remain about how these regulatory developments will be conducted, verified, and enforced in practice in a way that protects people's fundamental rights and is compatible with technical requirements. In particular, there is a lack of guidance on how companies and assessors should implement risk assessment and auditing mechanisms in line with the UNGPs, and how civil society and academia can most meaningfully engage around these processes.

The UN Human Rights B-Tech project, together with BSR, GNI, and Shift, helped develop several papers digesting and explaining how the international human rights and responsible business frameworks should guide approaches to risk management related to generative AI. More work is needed to understand how business and human rights practices can inform and bridge AI-focused risk assessments in the context of regulations like the DSA and the EU AI Act, and also to engage with the technical community on these implications.

Confidential - Not For Public Release

HIGHLY CONFIDENTIAL GOOG-HJCFG-00000724

This event, co-hosted by the Delegation of the European Union to the UN, the Global Network Initiative, the UN Human Rights B-Tech Project, and Humane Intelligence seeks to explore the following questions:

1. The landscape of AI regulation:

- What are key global trends with regard to regulation requiring tech companies to assess human rights risks?

2. Identifying suitable methodologies to assess AI products and services with regard to human rights risks:

- How can stakeholders (including engineers) encourage comparable AI risk assessment and auditing benchmarks?
- What might appropriate methodologies for AI auditing look like and what data is needed to perform accountable AI audits?

3. Requirements for effective oversight/enforcement and involvement of experts:

- What is the role of enforcing/supervisory mechanisms? (e.g. EU AI Office)
- How can civil society and academia most meaningfully engage around these processes?
- How can AI risk assessments and audits be used by companies and external stakeholders to ensure accountability and catalyse change?

Agenda

14.00: Opening remarks: **Ambassador Lotte Knudsen**, Head of EU Delegation to the UN in Geneva, **Peggy Hicks**, Director, Thematic Engagement, Special Procedures and Right to Development Division, United Nations Human Rights

14.15-15.10: Opening panel

- **Juha Heikkila**, Adviser for AI in the European Commission Directorate-General for Communications Networks, Content and Technology (CNECT)
- **Lene Wendland,** Chief Business and Human Rights, United Nations Human Rights
- **Mariana Valente**, Deputy Director Internet Lab Brazil/Professor of Law, University of St. Gallen, member on the Commission of Jurists for the AI Bill, Brazil
- **Alex Walden**, Global Head of Human Rights, Google
- Moderator: **Jason Pielemeier**, Executive Director of Global Network Initiative

Followed by Q&A

15.15-16.15: Table-top exercise including break-outs, facilitated by **Rumman Chowdhury**, CEO of Humane Intelligence

16.15 Closing remarks, followed by coffee, tea, and pastries

Confidential - Not For Public Release

HIGHLY CONFIDENTIAL GOOG-HJCFG-00000725

Exhibit 294

DSA RISK ASSESSMENT REPORT 2024

Confidential
28 August 2024
(Updated 2 October 2024)

CONTENTS

1. FOREWORD 1
2. EXECUTIVE SUMMARY 2
3. INTRODUCTION 4
4. RISK GOVERNANCE 5
5. AIGC, RECOMMENDER SYSTEMS AND RISKS TO USERS' PHYSICAL AND MENTAL WELL-BEING 10
6. ILLEGAL CONTENT: RISKS OF ILLEGAL HATE SPEECH CONTENT 15
7. ILLEGAL CONTENT: RISKS OF TERRORIST CONTENT 22
8. ILLEGAL CONTENT: RISKS OF INTELLECTUAL PROPERTY INFRINGING CONTENT 30
9. ILLEGAL CONTENT: RISKS OF CHILD SEXUAL ABUSE MATERIAL AND CHILD SEXUAL EXPLOITATION AND ABUSE 35
10. ILLEGAL CONTENT: RISKS OF GENDER-BASED VIOLENCE CONTENT 42
11. YOUTH SAFETY: RISKS RELATED TO AGE APPROPRIATE CONTENT AND ONLINE ENGAGEMENT 47
12. YOUTH SAFETY: RISKS RELATED TO AGE ASSURANCE 54
13. CIVIC INTEGRITY: RISKS TO ELECTIONS AND CIVIC INTEGRITY 59
14. CIVIC INTEGRITY: RISKS TO PUBLIC HEALTH FROM MEDICAL MISINFORMATION CONTENT 68
15. CIVIC INTEGRITY: RISKS TO PUBLIC SECURITY FROM HARMFUL MISINFORMATION CONTENT 73
16. FUNDAMENTAL RIGHTS: RISKS TO FUNDAMENTAL RIGHTS 79
ANNEX 1: KEY INFORMATION ABOUT TIKTOK 85
ANNEX 2: HOW TO USE THIS REPORT 87
ANNEX 3: SPECIALIST TEAMS 88
ANNEX 4: STAKEHOLDER ENGAGEMENT 93
ANNEX 5: OVERVIEW OF FACT-CHECKING PROGRAMME 111

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

1. FOREWORD

TikTok presents here its second annual systemic Risk Assessment Report within the framework of the Digital Services Act (**'DSA'**). TikTok remains steadfast in its mission to inspire creativity and bring joy whilst being a safe and welcoming community. Each day, millions of Europeans come to TikTok to find entertainment, education and fun.

The DSA's focus on transparency and systemic risk assessments aligns with our commitment to accountability and continuous improvement in managing risks. TikTok is dedicated to building a range of effective processes and policies to safeguard our community from the various content and conduct risks that are detailed in this Report. The challenges faced by online platforms are complex and ever evolving. To address these, TikTok strives to develop and support internal teams with expertise across systemic risk areas. Our efforts are further strengthened through close collaboration with regional and country-specific teams across Europe, combining specialised knowledge with local insights to enhance our safety initiatives. This local knowledge has been vital in tackling the significant volume of elections in the EU in 2024 and we are grateful to the European Commission for their engagement and guidance on this important topic.

TikTok has closely monitored existing and emerging risks over the last year, building on the insights gained from our first systemic risk assessment process. We have further developed our risk identification and mitigation strategies to enhance platform safety and security. Three key initiatives have been central to this effort: (1) investing further in stakeholder engagements to strengthen the ways in which we consider external perspectives to risk, which includes specific engagements in relation to the Israel-Hamas war; (2) addressing the known and potential risks associated with Artificial Intelligence, notably AI-generated content; and (3) ensuring we are engaging meaningfully and thoughtfully on the risks associated with our younger users aged between 13 and 17 years old. We expand on each of these areas in more detail below.

TikTok has strengthened its approach to external stakeholder engagement as part of our risk governance. In consideration of the text of Article 34 and Recital 90 DSA, we regularly engage with a wide range of independent experts, researchers, vulnerable groups and civil society organisations who inform our risk identification and mitigation design activities. We also do this through initiatives such as our European Safety Advisory Council, which is made up of diverse experts on systemic risks, and the TikTok Youth Council. The Youth Council comprises teenagers from around the world and provides a forum where we can hear directly from our younger users. We use these perspectives to inform changes we make to create the safest possible experience for our community. The Council's priorities for this year are on strengthening our approach to both teen well-being and inclusion on TikTok. We have included a summary of these important interactions at Annex 4 to this Report.

In last year's Risk Assessment Report, we identified emerging risks associated with AI-generated content and these risks have continued to evolve. TikTok has already proactively deployed processes to identify and mitigate such harmful AI-generated content, resulting in new and improved mitigations documented in this Report, and we are expecting to further strengthen our measures in the coming year.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014594

TikTok continues to prioritise the safety and well-being of its younger users, and has implemented new and updated policies and enforcement measures to support youth safety on the Platform. Key initiatives include the introduction of additional privacy settings, and expanded screen time interventions to promote responsible digital habits. TikTok has also launched a global multi-stakeholder initiative focused on age assurance in the digital environment. This initiative, involving industry partners and leading safety and privacy organisations, aims to create a holistic dialogue on age assurance complexities and establish working groups to address key issues, balancing child protection with safe online participation. TikTok remains committed to creating a safe and supportive environment for its younger users.

Our progress in the year since the implementation of the DSA has been significant, but we recognise that safety is an ongoing commitment. We are dedicated to monitoring our efforts and upscaling our mitigation strategies iteratively and based on the best available research. Looking forward, we welcome feedback on our approach from regulators, researchers and civic society, as such inputs are crucial to refining our risk management strategy.

Redacted

Redacted TikTok

2. EXECUTIVE SUMMARY

Introduction

This Report summarises the results of TikTok's second annual Systemic Risk Assessment within the framework of the DSA. TikTok submitted its first Systemic Risk Assessment Report to the European Commission in September 2023. This Report uses the terms '**Year 1**' to refer to the 2023 report and assessment, and '**Year 2**' to reflect this updated assessment of systemic risks, and any new or improved mitigations that have been implemented from September 2023 to August 2024.

Methodology

In Year 1, TikTok implemented a risk assessment framework to analyse the systemic risks present on its Platform. That framework scored each inherent and residual risk on a five-level scale, based on the effectiveness of existing mitigation policies, systems, and procedures. The framework guided decisions on when and how to apply further mitigations in a manner that is reasonable, proportionate, and effective.

Building on this foundation, in Year 2, TikTok strengthened this framework by incorporating elements of the ISO31000 risk scoring methodology. The Year 2 methodology examines the design and operating effectiveness of mitigations/controls to produce a comprehensive risk assessment and mitigation plan.

AIGC, Recommender Systems and risks to physical and mental well-being

TikTok has conducted a thorough risk assessment, which identified two priority factors that pose potential risks across many of its assessments, these being content generated by Artificial Intelligence

Technology ('**AIGC**') and the impact of recommender systems. TikTok also sets out its approach to analysing the potential serious negative consequences to users' physical and mental well-being.

Risk assessment results

TikTok has taken a diligent and cautious approach to the analysis of systemic risks that may arise from the design, functioning, use or mis-use of the Platform. TikTok's methodology remains substantially the same as last year. TikTok has classified the risks applicable to the Platform into 11 **Risk Modules** and then organised each Risk Module into 1 of 3 tiers to show how it prioritises its ongoing work to address those risks ('**Tiers**').

These Tiers reflect TikTok's current assessment of the priority level for each systemic Risk Module, considering existing policies, systems, and procedures for risk mitigation. This Tiering system is integral to TikTok's resource allocation. Special emphasis has been placed on risks affecting vulnerable groups and those risks more likely to evolve over the next 12 months. This overall approach ensures that TikTok is prepared to mitigate risks effectively and proportionately, maintaining a strong focus on community safety.

SUMMARY OF RISK ASSESSMENT RESULTS		
Tier 1 risks	Tier 2 risks	Tier 3 risks
• Risks of child sexual abuse material and child sexual exploitation • Risks related to age assurance • Risks related to age appropriate content and online well-being • Risks to elections and civic integrity • Risks of gender-based violence content	• Risks of terrorist content • Risks of illegal hate speech content • Risks to public security from harmful misinformation content • Risks to fundamental rights	• Risks of intellectual property infringing content • Risks to public health from medical misinformation

In Year 2, TikTok observed some changes in the risks to users of Medical Misinformation. This is due to its conclusion that the risk has now pivoted from content related to the Covid-19 pandemic to content related to a group of less widespread risks (See section 14). Medical Misinformation has therefore moved from a Tier 2 risk to a Tier 3 risk. TikTok also observed an increase in the potential for risks to fundamental rights as a result of the ever-evolving challenges of moderating complex hate speech content, driven by global conflicts but impacting content consumed in the EU. Risks to fundamental rights have therefore moved from a Tier 3 risk to a Tier 2 risk. All other risks remain in the same Tier, with the same emphasis of prioritisation as in Year 1. TikTok will continue to mature and improve its approach to the identification and mitigation of systemic risks where these can stem from the design, functioning or (mis)use of the Platform and its related systems, with particular regard to emerging risks.

3. INTRODUCTION

This DSA Risk Assessment and Mitigation Report (the **'Report'**) has been prepared in compliance with Article 42(4)(a),(b) and (e) of the DSA. This Report summarises the results of TikTok's second systemic risk assessment under Article 34 and the mitigation measures that have been put in place under Article 35 DSA. It has been prepared by TikTok Technology Limited (**'TikTok Ireland'**) in relation to the operation in the European Union (**'Europe'** or **'EU'**) of its online platform named TikTok (referred to as either **'TikTok'** or the **'Platform'** depending on the context within this Report), which has been designated as a Very Large Online Platform (**'VLOP'**) under the DSA. The DSA requires VLOPs to conduct systemic risk assessments at least once annually. This Report has been reviewed and approved by the board of directors of TikTok Ireland, following consultation with TikTok's head of the independent compliance function.

TikTok has enhanced its report structure and contents since its Year 1 Report. This Year 2 Report:

- Expands upon the risk identification and risk description from the Year 1 Report, with a dedicated section for an analysis of additional risks identified during Year 2;
- Includes an Annex detailing all relevant external stakeholder engagements and how they have led to risk identification or risk mitigations across each systemic risk in order to make observing compliance with Article 42(4)(e) as straightforward for the reader as possible;
- Addresses **Youth Safety** in two distinct sections to reflect the underlying risk assessment methodology (one for age appropriate content and online engagement; and a second for age assurance);
- Refers to '**Younger Users**' and 'Youth' rather than 'Minor(s)' as it did in Year 1. 'Younger Users' and 'Youth' refer to users aged 13-17. '**Underage Users**' refers to anyone accessing the Platform who is under the age of 13; and
- Includes a distinct **'Risk Environment'** section with an assessment of how intentional manipulation of the service may influence systemic risks in line with Article 34.2.

And also note:

- The section 'Key information about TikTok' can now be found in Annex 1 which sets out information about TikTok's content moderation systems, data related practices, advertising systems, recommender systems; and
- The Year 2 Report sets out the results of each Risk Module, grouped into topic-based risk categories (in the following order: illegal content, youth safety, civic integrity and fundamental rights), rather than grouping them by Tier-based prioritisation as TikTok did in Year 1.

This Report provides the results of TikTok's Year 2 systemic risk assessment through the following sections:

- TikTok's methodology and approach to the governance of systemic risks (see section 4);
- An analysis of the interplay of systemic risks and two horizontal priority areas, being risks related to AIGC and recommender systems (see section 5);

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014597

- An analysis of how TikTok has addressed its analysis of risk of 'serious negative consequences to [users'] physical and mental well-being', (see section 5); and
- The summarised results of TikTok's analysis of each systemic risk area, broken down (as it was in Year 1) thematically across illegal content, youth safety and well-being, civic integrity, and fundamental rights. The summary of each of these areas of systemic risk includes an assessment of the severity and probability of each risk. The Report also includes an assessment of the effectiveness (in reducing residual risk), reasonableness and proportionality of TikTok's mitigation measures. The Report incorporates identification of further mitigation actions for the year ahead (see sections 6-16).

This Report is supplemented by a **Stakeholder Engagement Report** in Annex 4, which summarises core stakeholder engagements that TikTok has conducted. These engagements are done to identify emerging risks and develop proportionate mitigation responses based on the best available information and scientific insights, in line with Recital 90 DSA.

4. RISK GOVERNANCE

A. INTERNAL RISK GOVERNANCE

TikTok recognises that effective risk management is critical to providing a safe and welcoming online platform. TikTok operates a risk management framework that emphasises a safety-first approach, integrates human rights standards, and fosters cross-functional collaboration.

TikTok strives to employ effective, reasonable, and proportionate risk mitigation measures tailored to identified systemic risks, whilst avoiding unnecessary restrictions to platform use or fundamental rights. Proactive risk management and detection strategies are also employed, even though general monitoring for illegal content is not required under the DSA.

Balancing user safety with other fundamental rights, TikTok follows the principle of proportionality, assessing risks at the intersection of safety, privacy, protection of youth, and freedom of expression and information. Acknowledging the complexity, TikTok is committed to achieving a fair balance for its users.

TikTok's policy-driven approach to risk management is reflected in its community guidelines, which establish a common code of conduct for all users and set out TikTok's content moderation practices ('**Community Guidelines**'). Continuous improvement is fostered through internal training and awareness programs, which drive well-informed decision-making and sound governance. Raising user awareness, particularly concerning disinformation campaigns, is also emphasised.

B. SYSTEMIC RISK ASSESSMENT METHODOLOGY

TikTok has built on and enhanced its Year 1 risk assessment methodology to inform its approach in Year 2. In TikTok's Year 1 Risk Assessment Report, TikTok organised the results of the risk assessments into three tiers. These tiers represented TikTok's assessment of the priority that a systemic risk category demands, having taken account of existing policies, systems and procedures for mitigating

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014598

the risk. Over the last year, these tiers have informed TikTok's view of when further mitigations are needed, and what kind of mitigation would be reasonable and proportionate.

The Tiering system allows TikTok to account for the level of inherent risk in certain systemic risk areas and therefore prepare for unforeseen events and the severity of potential harm to its user community, with a specific focus on vulnerable or marginalised groups. This ensures TikTok is resourced to put in place necessary controls and mitigations for a residual risk which is reasonable, proportionate, and effective.

TikTok has strengthened the detail of its risk scoring methodology (which is based on ISO31000 standards) since Year 1, as it has further leveraged subject matter expertise both internally and through external engagements, and considered new data sets in its risk scoring analysis. TikTok is ready to further develop this methodology should any guidance become available from the European Commission and once it has had the opportunity to consider how other platforms have approached this process.

The methodology comprises six key phases, all of which have been reviewed, and repeated for Year 2.

Phase 1	**Phase 2**	**Phase 3**	**Phase 4**	**Phase 5**	**Phase 6**
Systemic risk scoping and definition	Risk identification	Inherent risk analysis and assessment	Identification and assessment of mitigations	Calculation of residual risk and analysis of future improvements	Reporting, governance, and review

Phase 1: Systemic risk scoping and definition

TikTok has developed and maintained a systemic risk taxonomy that outlines the scope of the systemic risks that could stem from the design, functioning, use or mis-use of its Platform. In Year 1, TikTok developed a process for scoping and defining risks based on an analysis of three key elements:

- The four primary systemic risk categories, as described in the DSA;
- A legal and textual interpretation of key DSA terms, other relevant EU law and jurisprudence, and relevant international legal instruments; and
- A functional interpretation made by TikTok's internal teams.

In Year 1, this process informed the development of a systemic risk taxonomy that included twelve distinct Risk Modules, grouped under four primary categories. In Year 2, following careful review for any developments that would require a change to the taxonomy, TikTok decided that it remained appropriate to use the Year 1 taxonomy without amendment. Therefore, as in Year 1, the four primary categories are '**Illegal Content'**, '**Youth Safety'**, **'Civic Integrity'**, and **'Fundamental Rights**', which are assessed in 12 distinct modules to ensure a comprehensive analysis of the risks and mitigations (**'Age Appropriate Content'** and **'Online Engagement'** have been summarised together in this Report due to their overlapping risks and corresponding mitigations, and '**Age Assurance'** is reported on separately).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014599

Phase 2: Risk identification

Phase 2 involves identifying the specific ways in which each risk area may manifest on TikTok. TikTok reviewed the risks identified in Year 1 and confirmed that they remain applicable in Year 2 and updated and improved the detail and specificity of risk identification within each module in Year 2. TikTok also identified any emerging or new risks.

This risk identification is informed by 3 key processes:

- Consideration of internal sources (e.g. data insights from content moderation practices);
- Consideration of external sources (e.g. stakeholder engagement, law enforcement reports, and credible independent research); and
- An assessment of potential emerging risks (e.g. proactive risk assessment, internal research, stakeholder engagement).

In identifying how the systemic risk may manifest on the Platform, TikTok assesses how it is influenced by the factors laid out in Article 34(2) DSA and other relevant recitals. Article 34(2) instructs platforms to assess 'the design of recommender systems and any other relevant algorithmic system; content moderation systems; applicable terms and conditions (policies) and their enforcement; systems for selecting and presenting advertisements ('ads'); [and] data related practices.' Where the risks are localised or there are linguistic differences, TikTok takes these into consideration when assessing risk and corresponding mitigations.

In the Year 2 Risk Assessment, TikTok has also set a heightened focus on two horizontal priority areas that are of a cross-cutting nature across several risk areas. These are:

- The impact of an increased prevalence of AIGC; and
- The impact of recommender systems on the content that users consume.

TikTok has examined the 'potential for serious negative consequences to users' physical and mental well-being' (both of adults and minors), which is specified in Art. 34(1)(d), in the process of risk identification in all Risk Modules. A range of such risks were detected in relation to all Risk Modules, apart from Risks of Intellectual Property Infringing Content.

TikTok's selection of the priority considerations has been informed by internal risk detection analysis and an analysis of priorities expressed by stakeholders, including the European Commission.

Phase 3: Inherent risk analysis and assessment

Within the assessment of inherent risk, TikTok has considered severity and probability for each Risk Module, prior to considering the impact of any implemented mitigation measures.

To assess inherent risk, severity and probability are quantified and weighted before being multiplied against each other. To quantify each factor:

- **Severity** is quantified by assessing the scope (i.e. number of users or affected persons in Europe), scale (ranging from individual impacts to consequences to the broader society), the nature of potential impact (physical, psychological or otherwise), and impacted demographics (which may include vulnerable groups). This calculation of severity is then weighted to ensure

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014600

that severity is accurately reflected and that lower scores in one or two factors do not overshadow high scores elsewhere; and

- **Probability** is quantified by assessing likelihood and historical volume. Likelihood is assessed based on factors such as whether (and how much) the risk has already been observed and whether it is expected to occur in the future. Historic volume is determined based on an assessment of historic occurrence on the Platform. Probability is therefore largely a measure of proxies; given that controls are in place, the probability of events occurring without any controls in place cannot be directly measured.

The calculated scores for severity and probability are multiplied to produce an inherent risk score for each Risk Module on a scale which comprises - as it did in Year 1 - 1 (very low), 2 (low), 3 (moderate), 4 (material) and 5 (high).

Phase 4: Identification and assessment of mitigations

TikTok has identified the mitigation measures that it has implemented relevant to each Risk Module, with reference to the types of mitigations set out in Art. 35(1)(a)-(k) DSA, noting any other additional mitigations that have been implemented in Year 2. Mitigation measures (i.e. controls) are assessed for their design and operating effectiveness. An analysis of the effectiveness of the mitigation is then used to calculate residual risk.

Phase 5: Calculation of residual risk and analysis of future improvements

Residual risk represents the amount of risk that remains for each risk area after the impact of relevant controls/mitigations on reducing inherent risk is considered. It is calculated by evaluating inherent risk against the effectiveness of the mitigations for each Risk Module to produce a residual risk for each Risk Module on a scale which comprises 1 (very low), 2 (low), 3 (moderate), 4 (material) and 5 (high). This analysis generates areas for the enhancement of mitigations, which form the work plan for future improvements.

Phase 6: Reporting, governance, and review

Each Risk Module is then reviewed and approved by subject matter experts and leaders to ensure robust governance and strong accountability, as follows:

1. The Risk Module is first submitted to a specialist Risk Assessment Review Group for analysis and assessment. The Risk Assessment Review Groups' process, including the Compliance function's input, is designed to ensure that all risks referred to in Article 34 are identified and properly reported on;
2. The results are then reported to the Online Safety Oversight Committee, TikTok's steering group of cross-functional leaders which has been in place throughout its DSA programme; and
3. Finally, the results are submitted to the Board of Directors of TikTok Technology Ireland for review and approval. This includes a decision from the Board on the prioritisation Tier of each risk.

C. RISK ENVIRONMENT

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014601

TikTok is one of a number of online platforms whose users face similar, but not identical risks. These risks often transcend the boundaries of individual platforms and involve complex interactions between online experiences and real-world events, facilitating personal expression and the receiving and imparting of information and ideas. This is a central value of the EU, a fundamental right, and a core pillar of European democracy and social life. Each platform is expected to operate on its own analysis of the appropriate balance between safety and fundamental rights for its users. TikTok takes an approach that prioritises caution in areas where risks are more likely to impact real-world harms, or where they may be more likely to impact Younger Users.

TikTok strives to only host content within the bounds of relevant laws and the rules it imposes on itself and its users. TikTok therefore aims, through its use of automated and human moderation measures, to proactively identify and remove content before it is seen by any user. However, operating within such complexity and considering the inherent risks of human expression, it is not possible to eliminate all risks, as mitigating one risk often increases another (e.g. severely restricting hate speech would magnify risks to freedom of expression). Therefore, TikTok aims for a balanced and proportionate approach to reducing risks continuously, in line with Recital 86 DSA.

While TikTok is not specifically aimed at minors or predominantly used by them, TikTok makes it a priority to address how risks may impact these users, please see sections 11 and 12 of this Report.

TikTok has a comprehensive crisis management plan to address unforeseen challenges. TikTok maintains a dedicated incident management team to address urgent issues and to contain and minimise harm. Additionally, TikTok is committed to learning from past incidents, adapting its strategies, and fortifying its Platform against future risks.

D. WHAT DOES TIKTOK DO TO COMBAT INTENTIONAL MANIPULATION OF THE SERVICE?

TikTok aims to foster an environment where genuine interactions and content thrive. However, bad actors may attempt to manipulate the service through deceptive behaviours like account impersonation, spam activity, and fake engagement. This can range from individual behaviour, such as creating a fake account, to sophisticated and coordinated influence operations.

To combat these issues, TikTok prohibits - and works to disrupt - intentional manipulation and inauthentic use of the Platform. This involves detecting and removing deceptive accounts by assessing technical signals related to user registration, account traffic patterns, engagement behaviours, and content volume. TikTok's verified account badge helps protect users by providing a reliable indication of high-profile accounts.

Expert teams at TikTok focus on detecting, investigating, and disrupting covert influence operations. They analyse behaviour and account linkages, monitor cross-platform threats, use internal detection signals, and manually investigate reports. Suspicious accounts and behaviours are automatically removed, and inauthentic engagement signals are ignored in estimating account or video popularity. TikTok reports up to monthly on the number of accounts actioned under its policies against covert influence operations in its public transparency centre.

TikTok recognises that covert influence operations will continue to evolve and that networks may attempt to re-establish a presence on the Platform. As new deceptive behaviours emerge, TikTok is committed

to evolving its response, strengthening enforcement capabilities, and publishing its findings for scrutiny and more effective cross-platform risk mitigation.

E. STAKEHOLDER ENGAGEMENTS

TikTok has actively engaged with a broad spectrum of stakeholders to strengthen its risk assessment and mitigation strategies under the DSA. This proactive approach is central to TikTok's commitment to user safety and the protection of fundamental rights on its Platform. Over the past year, TikTok has consulted with a diverse range of experts, including members of its European Safety Advisory Council **(ESAC)**, policymakers, civil society organisations, and content creators. Through close collaboration with stakeholders across Member States, TikTok can better understand and respond to local contexts, ensuring that its safety initiatives are not only globally robust but also finely tuned to address specific regional concerns. These partnerships enable TikTok to combine its existing Trust and Safety teams' expertise with further local knowledge, creating more effective and targeted risk mitigations, including campaigns that support a safer online experience for different communities. These engagements have directly informed TikTok's development of mitigation measures addressing systemic risks and the emerging risks associated with AIGC.

Annex 4 highlights TikTok's comprehensive strategy for engaging with stakeholders, which has been integral to refining its safety policies and operations. It details how TikTok's collaborations with external experts have directly influenced the Platform's approach to identifying and mitigating risks. These efforts include targeted consultations on emerging threats, the creation of specialised advisory councils, and ongoing dialogues with civil society. By leveraging this extensive network of partnerships, TikTok has been able to develop and implement measures that are reasonable, proportionate, and effective in addressing the complex challenges associated with hosting user generated content (**'UGC'**).

F. HOW TIKTOK COLLABORATES WITH DSA TRUSTED FLAGGERS

Trusted Flaggers have a key role to play in risk identification and TikTok welcomes the designations of three organisations as Trusted Flaggers under Article 22 DSA in the past few months. In the Year 1 Report, TikTok stated as a mitigation measure that it would 'engage with organisations when they are newly designated as 'trusted flaggers' by EU member states and undertake outreach to onboard such entities to ensure efficient and priority processing of such reports via TikTok's dedicated channels as part of the Trusted Flagger Engagement Strategy'. As a cross-risk year 2 mitigation, TikTok remains committed to this and continues to be ready to engage with such partners as they are designated. In the meantime, TikTok continues to take action on reports received from its Community Partner Channel and trusted subject matter experts.

5. AIGC, RECOMMENDER SYSTEMS AND RISKS TO USERS' PHYSICAL AND MENTAL WELL-BEING

Introduction

Through its risk assessment process, TikTok has identified that AIGC and recommender systems are two critical factors that pose potential risks across several systemic Risk Modules. In response, TikTok has - both pre-dating the DSA entering into force and since that time - adapted and implemented a

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014603

series of mitigation strategies to address these risks comprehensively across the Platform. These measures are designed to safeguard the user experience, and ensure the Platform's resilience against emerging threats in these areas. Since these factors influence specific Risk Modules in unique ways, they are addressed in relevant sections throughout the Report.

TikTok acknowledges the importance of thoroughly assessing the risk of possible 'serious negative consequences to users' mental and physical well-being'[1]. TikTok has evaluated these risks to both Younger Users and other users when identifying risks in all Risk Modules. TikTok has identified a range of such risks in Risk Modules concerning Illegal Content, Youth Safety and Civic Integrity and evaluated how such risks may arise through use of the Platform.

A. THE IMPACT OF INCREASED USE OF AIGC

Definition and cross-module risks

In its Year 1 Report, TikTok identified potential emerging risks associated with AIGC in the online platform ecosystem. TikTok assessed that potential risks could arise through the sharing and dissemination of deceptive - but extremely realistic - images, videos and audio, potentially making it more difficult to distinguish between fact and fiction on the Platform. TikTok also foresaw potential new opportunities for bad actors to create and disseminate illegal content (including image-based abuse). Since the last risk assessment, these risks have materialised online, as there has been significant growth in the number of providers of tools to produce AIGC and a rapid increase in the amount of AIGC created.

As AIGC tools can be used to create content that appears just the same as UGC, the use of such tools creates potential risks to the integrity and reliability of content hosted on TikTok. **'Deep fakes'** - extremely realistic photos, videos or voices of real people - have emerged as a particular area of concern, given their risk of harm to depicted individuals. Additionally, AIGC technologies can be used to facilitate the creation of illegal content (e.g. synthetic Child Sexual Abuse Material, '**CSAM**') and image-based sexual abuse.

AIGC tools are now widely available to create manipulated content at low cost and high speed. AIGC is not inherently harmful, but the increasing accessibility of these tools and the nascence of best practice mitigation strategies within them magnifies the chance of harmful content being created and disseminated.

Cross-risk mitigations

In Year 2, TikTok has continued to implement measures across both its UGC and its advertising features to mitigate potential risks associated with AIGC.

Regarding UGC, TikTok's Community Guidelines prohibit any AIGC that would violate its regular policies in respect of illegal and violative content. In its update to the Community Guidelines earlier this year, TikTok made changes to reflect recent developments in AIGC, labelling and to reinforce existing policies. For example, AIGC uploaded by users that portrays CSAM is prohibited in the same way that any CSAM is prohibited, regardless of how it is created. TikTok has also developed specific policies regarding the creation and dissemination of AIGC, including:

[1] Article 34(1)(d) DSA.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014604

- Requiring users to disclose content that is either completely generated or significantly edited by AI and contains realistic-appearing scenes or people;
- Prohibiting AIGC that depicts realistic-appearing people under the age of 18;
- Prohibiting AIGC that depicts the likeness of adult private figures without their permission;
- Prohibiting misleading AIGC that appears to come from an authoritative source;
- Prohibiting misleading AIGC that depicts a crisis event; and
- Prohibiting misleading AIGC that depicts a public figure who is being degraded or harassed, engaging in criminal or anti-social behaviour, or being politically endorsed or condemned.

With regard to TikTok's advertising features, TikTok's regular UGC policies apply, alongside additional advertising-specific rules. TikTok has developed specific AIGC Advertising Policies, which prohibit ads:

- Containing AIGC or suspected AIGC in relation to Crisis Events or Elections;
- Containing speech fabricated deep fakes;
- Using AIGC or suspected AIGC to display events that did not happen, or mislead users on true/real events;
- Using AIGC or suspected AIGC to make it look like it comes from an authoritative source;
- That contain the likeness of adult private figures without their permission;
- That depict public figures in deceptive ways (e.g. AIGC falsely depicting an endorsement by a public figure is banned);
- Featuring AIGC political figures; and
- Which TikTok knows or suspects features AIGC imagery depicting people aged 18 or under.

TikTok is prepared to further iterate and improve these policies as AIGC trends and risks evolve, and enforcement capabilities evolve.

For the purposes of detecting and enforcing against the use of AIGC to create violative content in either a UGC or ads context, TikTok has implemented tools for users and advertisers to voluntarily label their content as 'AI-generated'. TikTok was the first platform to launch an AIGC labelling tool, which over 37 million creators have used since September 2023 worldwide. From Q3 2023 to Q2 2024, 1,668,376 unique users in the EU have labelled their content with the 'Creator labelled as AI-generated' label, accounting for 5,864,626 unique videos. Additionally, any AIGC that is made with TikTok's AI effects is automatically labelled. If users believe that content seen is in fact undisclosed AIGC, they can report it to TikTok, who may then remove the content if it is otherwise violative of its Community Guidelines regarding AIGC.

TikTok also works with industry groups and civil society to detect undisclosed AIGC. TikTok participates in the Coalition for Content Provenance and Authenticity ('**C2PA**'), an open technical standard and content provenance solution that can provide information in a piece of content's metadata about its origins and whether AIGC models were used to create or edit it. This technology helps TikTok to detect AIGC at scale when content with C2PA metadata is uploaded to the Platform. Through its use of C2PA, TikTok has been the first video sharing platform to automatically label AIGC that has been made on other platforms. From Q3 2023 to Q2 2024, 4,286,027 videos on TikTok were auto-labelled with the AIGC tag of 'AI-generated' in the EU.

TikTok is also developing improved internal AIGC detection tools to enhance the scale, reach and accuracy of its enforcement against violative AIGC. This includes two automated models currently under development:

1. A 'visual model' to detect whether a relevant image or video is AIGC; and
2. An 'audio model' to detect whether the audio was created is AIGC.

Finally, TikTok's capacity to effectively detect AIGC is reinforced by external partner reports, internal teams who identify instances of violative use of AIGC, manual reviews of viral content, and targeted proactive sweeps[2] of keywords, URLs, or hashtags associated with high-risk content. TikTok's approach has been informed by research conducted by the Trust & Safety Product Policy and Issue Policy teams on academic, industry, and advocacy literature. TikTok has also consulted external bodies and subject matter experts, such as the Partnership on AI.[3]

B. RECOMMENDER SYSTEMS

Definition and cross-module risks

The main feature on TikTok that uses a recommender system is the For You Feed (**'FYF').**[4] This feature is the primary means through which users consume video content on the Platform. The FYF uses a personalised recommendation system to help each user discover content, creators, and topics, while ensuring that content is interesting and relevant to each user. In determining what is recommended, the FYF's recommender system takes account of user interactions, content information, and user information.

Concentrated content[5] is a relevant risk that has been widely documented by civil society.[6] As recommender systems are designed to offer content based on a user's interests and historical engagement, they might inadvertently present users with a narrow range of content for extended periods of time. Certain types of concentrated content, though not violative of TikTok's Community Guidelines, may cause harm by inadvertently reinforcing a negative personal experience for some viewers. For example, there may be an impact to mental well-being, particularly for Younger Users, associated with concentrated content relating to extreme dieting and body-image-related content.

Given that the type of content that may have a potential negative impact when viewed in high concentrations is not illegal or violative, it is challenging to assess and detect the impact or the appropriate mitigation measures. There are also risks to freedom of expression associated with acting against concentrated content that has a negative impact but which, in isolation, does not violate TikTok's Community Guidelines (e.g. extreme dieting or fitness).

---

[2] Proactive sweeps refer to the active scanning or reviewing of content to identify and remove violative content. It is proactive because it involves the initiative to find and address content rather than responding to reports.

[3] A non-profit organisation which brings together companies and other stakeholders to establish common codes of practice across the industry.

[4] TikTok has several other features which are powered by its recommender system, as outlined here:https://support.tiktok.com/en/using-tiktok/exploring-videos/how-tiktok-recommends-content.

[5] Note that this was referred to as 'filter bubbles' in the Year 1 Report.

[6] https://policyreview.info/articles/analysis/recommender-systems-and-amplification-extremist-content

Cross-risk mitigations

TikTok's enforcement strategies mitigate against the risk of amplifying illegal and harmful content through recommender systems. First, TikTok proactively enforces its Community Guidelines against illegal and harmful content, preventing the most egregious content from appearing on users' FYFs. Second, TikTok maintains content eligibility standards for FYF that prioritise safety and are informed by the diversity of community and cultural norms. For example, unverified information awaiting fact-checker review is ineligible for FYF. Similarly, TikTok ensures that content that directly attacks individuals or groups with protected attributes is not eligible to appear in its FYF. Third, TikTok integrates age-appropriate design into FYF; as content created by anyone under 16 years old is ineligible for recommendation and its content classification system rates UGC based on maturity and prevents inappropriate themes from being recommended to Younger Users. Fourth, to mitigate additional risks associated with viral content, TikTok manually reviews all content that reaches a certain level of popularity (by monitoring the volume of video views).

To mitigate risks from concentrated content, TikTok uses dispersion techniques in its FYF. This involves using machine learning models to avoid recommending a series of similar videos on themes that do not violate TikTok's Community Guidelines but are potentially problematic if viewed repeatedly.

TikTok also offers users tools to empower users to influence what they see on their FYF. This includes tools to understand why videos have been recommended, to filter out certain keywords or hashtags to stop seeing certain content, and to 'refresh' their FYF to view a new set of popular content as if they were a new user. At the end of H1 2024, 2,109,240 users had actively filtered hashtags in the EU.

C. POSSIBLE SERIOUS NEGATIVE CONSEQUENCES TO USERS' PHYSICAL AND MENTAL WELL-BEING

TikTok sets out below how it has approached the systemic risk of 'serious negative consequences to the person's physical and mental wellbeing (Art. 34(1(d) DSA). Given the absence of consensus among health professionals and academics on how social media use may adversely affect user health, TikTok engages experts and monitors associated risks. TikTok observes that these risks arise both from: (1) the risks from the consumption of violative or illegal content; and/or (2) as a result of use of the Platform itself, even though the content being consumed may not in itself be harmful. We set out below further details of how TikTok has analysed these risks.

Risks from the consumption of violative or illegal content

TikTok has addressed such impacts for all its users throughout this assessment, with reference to the areas of systemic risk in the Risk Module to which they relate. Each Risk Module has been categorised to ensure that the risks to mental and/or physical health are clearly understood and mapped. For example, risks of mental harm, for example due to exposure to radicalising Hate Speech content, are considered in the Risk Module concerning Hate Speech. Risks of physical harm, for example relying on false medical information, are discussed in the Medical Misinformation Risk Module. TikTok also applies extra measures to protect Younger Users from the risks of such content which are considered in Section 11 below.

Risks arising from use of the Platform itself

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014607

While TikTok does not believe that use of platforms (in general or of TikTok in particular), is automatically harmful in and of itself, it recognises that some users (in particular Younger Users) could at times and in specific circumstances face serious negative mental and physical health impacts as a result of the way in which they use such platforms.

The nature and causes of such impacts are complex and highly context-dependent, and frequently interact with (or arise from) other areas of systemic risk. While these risks apply to users generally, they are particularly impactful when it comes to Younger Users. TikTok has therefore decided to address such impacts in the Risk Modules on Younger Users. For example, TikTok's analysis of the potential impacts of increased screen time and social media use for Younger Users, which stems from its awareness of the particular challenges youth can face around self-regulation and impulse control, forms part of the Online Engagement Risk Module. TikTok sets out in that Module which mitigations also apply, differentiating between those available to all users and which apply to Younger Users. For example, TikTok offers screentime controls to both adults and Younger Users in order to facilitate the management of time spent on the platform. These controls are turned on by default for Younger Users and also form part of the suite of controls in TikTok's Family Pairing feature. TikTok recognises that the impact of screentime and social media use is a rapidly developing area of research. In particular, TikTok monitors the study of potential links between high levels of screentime and social media use and negative mental health outcomes. In parallel, TikTok is continuing to study potential risk factors in this area and is closely monitoring research developments in order to inform its analysis (including, in the context of Digital Services Act compliance, the European Commission's assessment of this topic).

As its understanding evolves, TikTok's assessment of the risks relating to physical and mental wellbeing will also evolve in order to ensure that appropriate steps are taken to continue to provide a safe environment for all TikTok users.

6. ILLEGAL CONTENT: RISKS OF ILLEGAL HATE SPEECH CONTENT

1. Description of the risk:

a. Description of the risk from Year 1

- TikTok understands the term **'Hate Speech'** in a manner consistent with EU and EU member state laws, in particular, having regard to EU Framework Decision 2008/913/JHA, Article 1, in respect of offences concerning racism and xenophobia (i.e., against a group of persons or a member of such a group defined by reference to sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation);
- The risks associated with Hate Speech may include users attempting to share or disseminate the following content including through video, livestream, comments and in profile information on the Platform:
    - Content claiming individuals or groups with protected attributes are physically, mentally, or morally inferior or referring to them as criminals, animals, inanimate objects, or other non-human entities;

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014608

- Content promoting or justifying violence, exclusion, segregation, or discrimination against them;
- Content that includes the use of slurs against others (and not the user themselves);
- Content that targets transgender or non-binary individuals through misgendering or deadnaming; or
- Content that depicts harm inflicted upon an individual or a group on the basis of a protected attribute.

- TikTok also appreciates that Hate Speech is in direct violation of the principles of liberty, democracy, respect for human rights and fundamental freedoms and, specifically, the right to non-discrimination as outlined in Article 21 of the EU Charter of Fundamental Rights (the '**Charter**').

b. Risk Identification: changes to the risk profile since Year 1

In its Year 1 Risk Assessment, TikTok classified Hate Speech as a **Tier 2 priority**.

Summary of risks identified in Year 1

TikTok identified the following risks related to Hate Speech in its Year 1 Risk Assessment:

- **Sharing or dissemination of Hate Speech**: The risk of users attempting to upload, stream or otherwise disseminate illegal hate speech on the Platform;
- **Content moderation systems**: The risk that content moderation systems may not keep pace with rapidly evolving alternative vocabulary can make hate speech difficult to effectively detect and remove from the Platform. However, TikTok notes that this risk must be balanced with the risk that over enforcement may create risks of harm to freedom of expression;
- **Ads:** The risk of advertisers attempting to disseminate Hate Speech on the Platform; and
- **Regional and linguistic complexities**: The risk that Hate Speech can be highly localised in terms of language and region and so, when combined with the risk of a rapidly evolving vocabulary, creates risks to the effectiveness of TikTok's policy coverage and moderation efforts.

Additional risks identified in Year 2

In Year 2, TikTok identified the following additional risk areas, which it considers to be reflective of the dynamic and evolving nature of Hate Speech:

- **Evolving Hate Speech content-related risks:** New risks related to trends in Hate Speech online, particularly marginalising speech, or content that indirectly demeans protected groups, such as anti-migrant hate, transphobic, and homophobic content. In the wake of the Israel-Hamas war, there has been a global increase in antisemitism, Islamophobia, and anti-Arab hate online. As of April 2024, TikTok has removed more than 3.1 million videos and suspended over 140,000 livestreams in Israel and Palestine for violating Community Guidelines, including content related to promoting Hamas, hate speech, violent extremism, and misinformation;
- **Risk of AIGC containing slurs or promoting hateful ideologies:** AIGC poses significant challenges to current Hate Speech detection systems. Certain third-party AI tools may be used to generate content with slurs or hateful ideologies that are subtly altered to evade detection by both human and automated moderation systems. This includes fake videos, audio and memes that spread Hate Speech or extremist content under the guise of humour, often using coded

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

language or symbols. For more detailed information on TikTok's approach to mitigating risks associated with AIGC, please refer to *Priority Considerations Across all Modules*.

c. Inherent Risk in Year 2

For a detailed analysis on how TikTok assessed the baseline severity and probability for Hate Speech in Year 1, please refer to TikTok's Year 1 Systemic Risk Assessment Report.

In Year 2, TikTok assesses the overall severity of Hate Speech to be 'Material'. TikTok attributes this to intensified global conflicts (e.g. Russia and Ukraine and Israel and Hamas) and challenges associated with potential uses of AIGC, which have amplified the potential for Hate Speech content to adversely affect the individuals or groups it targets. TikTok assesses the probability of Hate Speech to be 'Likely' in Year 2. Taking this into account TikTok has assessed the inherent risk (that is, the risk without any mitigations in place) of Hate Speech, being shared on the Platform in Year 2 to be 'Material'.

2. Mitigation Measures:

For detailed information on TikTok's system of mitigations/controls (under DSA Art. 35 (1)(a)-(k)) that applied when its designation as a VLOP came into force, please refer to the Year 1 Risk Assessment Report. That Year 1 Report identified further mitigation improvements, progress on which (and any other such mitigations) is listed below. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk monitoring and detection processes, and extensive consultations with external stakeholders.

Implementation of Additional Mitigation Effectiveness Measures in Year 2	
MEASURES IN ACCORDANCE WITH DSA ART. 35(1) A-K	**DESCRIPTION**
(b) Adaptation of terms and conditions	In addition to the mitigation effectiveness improvements anticipated in the Year 1 Report, TikTok has launched and enforced new Community Guidelines on marginalising speech in order to prevent the dissemination of content that indirectly disparages protected groups. TikTok has designated content that uses stereotypes, insinuation, or indirect statements that may implicitly demean protected groups as **'Feed Ineligible'** to strike a balance between mitigating harmful content and allowing free speech This means the content may remain accessible but will not appear on the FYF. Enforcement of this policy is driven by internal teams with subject matter expertise. In its Community Guidelines on Hate Speech and Hateful Behaviours (which includes Hate Speech & Hateful Ideologies, Slurs, Negative Stereotypes and Generalisations, and Disparaging Religion), TikTok has included more examples of protected attributes and groups and has reviewed the examples of prohibited behaviour provided to ensure they are balanced across protected groups.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014610

(c) Adaptation of content moderation processes	In addition to the mitigation effectiveness improvements anticipated in the Year 1 Report, TikTok has delivered a new programme to ensure that all content moderators complete training on anti-bias awareness. This training aims to increase awareness of potential biases in decision-making and equip moderators to recognise and mitigate bias effectively, and was developed with input from anti-bias experts and reviewed by advocacy groups. TikTok also expanded its specialised process for Hate Speech video moderation to enhance enforcement following updates to the Hate Speech policy. This launch includes hate policy specialist moderators who possess relevant language skills and cultural context, who are equipped with moderation technology providing full watch and listen capabilities. This expansion aims to improve overall enforcement quality. The new specialised content moderation queues dedicated to hate-related content were piloted in April 2024 and are now in operation in a subset of European markets - France, Germany, and Italy. TikTok will further extend this to cover Spain, Netherlands and Sweden throughout the rest of 2024. Finally, at the end of 2023 TikTok developed a global policy reminder[7] related to Antisemitism and Islamophobia. The complexity and evolving nature of Hate Speech may result in new trends, especially in the wake of a conflict situation. This policy reminder ensures that moderators are able to identify and remove Hate Speech content in the wake of the Israel-Hamas war.
(d) Adaptation of algorithmic systems	In addition to the mitigation effectiveness improvements anticipated in the Year 1 Report, TikTok has launched two additional automated moderation models to detect and designate Hate Speech content as ineligible for the FYF, which means that the content will not appear in the FYF but is discoverable if a user searches for it specifically, to correspond to the new Community Guidelines on Marginalising Speech.
(f) Reinforcing risk detection measures	In its Year 1 Report TikTok stated that it would 'expand its collaboration with external partners in order to develop enhanced intelligence-gathering capacities in relation to Hate Speech. This should assist in early detection of emerging and new forms of Hate Speech.' For example, since January 2023, TikTok has partnered with Protection Group International to identify and proactively mitigate extreme-risk areas, processing about 35 intelligence reports on violent behaviour and hate globally.

[7] Policy Reminders are documents with refreshed and updated policy information for moderators. These are usually created in response to a crisis event (such as the Israel-Hamas war) or in preparation for an event where TikTok may anticipate specific policy issues (such as Hate Speech before the Euros football). They do not change the policy but rather provide detailed guidance in the instance it is needed.

	Trust & Safety teams are dedicated to proactive forecasting and detection of risks, including Hate Speech, leveraging regional and subject matter experts and the Global Major Events Calendar, news alerts, in depth trend reports, and social listening. Every year, TikTok convenes cross-functional teams to identify and mitigate Hate Speech related risks that may arise around events such as Pride, Black History Month, Holocaust Remembrance, and sports events such as the Six Nations.
(g) Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21	As stated in Year 1, TikTok is a committed signatory of the EU Code of Conduct on Countering Illegal Hate Speech Online (the '**Hate Speech Code**'), and will continue its engagement with the European Commission and key stakeholders as part of this important initiative to combat online Hate Speech. TikTok remains ready to engage with EU 'trusted flaggers' as they continue to be designated (read section 5) and in the meantime, TikTok continues to receive reports of Hate Speech from Community Partners, composed of advocacy groups and experts.
(h) Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct and the crisis protocols referred to in Articles 45 and 48 respectively	As noted in Year 1, TikTok has continued to participate as a signatory to the Hate Speech Code. In collaboration with the European Commission and other participating platforms, TikTok has been actively involved in the revision and improvement of the Hate Speech Code since 2020. TikTok has committed to sign up to the revised Hate Speech Code, which is intended to be adopted under Article 45 DSA later this year. As part of the new revised Hate Speech Code, TikTok has committed to significant new obligations. These include a commitment to review 50% of valid illegal Hate Speech user notices within 24 hours, demonstrating TikTok's dedication to swiftly addressing illegal content on the Platform. For detailed information on how TikTok has engaged with external stakeholders to inform its approach to identifying and mitigating the risk of Hate Speech Content, please refer to Annex 4.

(i) Taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information	In its Year 1 Report, TikTok stated that it would 'build on the success of media literacy campaigns, such as the #SwipeOutHate campaigns, and continue such initiatives and consider further media literacy measures to generate awareness of issues relating to illegal content and of the available safety tools.' TikTok has signed a new partnership with the Six Nations rugby tournament to continue to run Swipe Out Hate campaigns including for the 2024 editions of the tournament. TikTok ran these campaigns to educate TikTok's users that TikTok does not tolerate hate speech and encourage its users to report hateful content during both the men's and women's Six Nations tournaments. TikTok has continued to deliver such initiatives and media literacy measures to generate awareness of issues relating to illegal content and of the available safety tools. TikTok has reviewed its communication to users when they violate Hate Speech policies, to help users better understand when and why content represents a Hate Speech violation. For detailed information on how TikTok has engaged with external stakeholders to inform its approach to identifying and mitigating the risk of Hate Speech Content, please refer to Annex 4.

3. Residual Risk in Year 2:

Following its assessment of the effectiveness, reasonableness and proportionality of relevant mitigations, TikTok has assessed the residual risk of Hate Speech to be Moderate in Year 2. TikTok commits to continuing to expand and invest in mitigation measures to identify and mitigate Hate Speech on the Platform.

This assessment was conducted on the basis of qualitative and quantitative insights, including data on TikTok's enforcement of its Community Guidelines in the EU for the period of Q3 2023 to Q2 2024. Under the Community Guidelines policy of 'Hate Speech & Hateful Behaviour', TikTok removed 1,083,469 total videos in the EU, with 898,809 detected and removed proactively, and 616,327 removed without any views. This high proportion of proactive removal suggests that TikTok's content moderation systems remain effective in detecting and removing Hate Speech on the Platform, helping to mitigate the systemic risk.[8]

This assessment of the effectiveness of TikTok's proactive detection systems is further informed by the following data on user reports, which serve as a useful indication of the volume of Hate Speech that is not proactively detected and removed by TikTok. User reports accounted for a minority of Hate Speech

[8] The percentage of video removals with zero views decreased throughout 2023 across all policy titles as a result of how TikTok prioritises content for human reviewers. TikTok's approach aims to minimise views of violative content, prioritise expeditious removal of especially egregious content (e.g. CSAM and violent extremism), and ensure consistency and fairness in enforcement. Other indicators such as total content removals reported may be substantively different to Year 1 as TikTok has enhanced its data integration practices by developing CG policy titles that are more specific to systemic risks, localising data to the EU, and regularising an annual reporting period.

detected and removed by TikTok, at 184,660 of the 1,083,469 total videos detected and removed under 'Hate Speech & Behaviour'. TikTok remains committed to continuously working to improve and iterate its proactive detection capabilities, while noting that Hate Speech is a rapidly evolving risk area, with nuance across regions and languages.

TikTok's user reporting tools are an important component of its content moderation process, ensuring that Hate Speech is detected and removed, including new and evolving forms of Hate Speech. The data demonstrates that when TikTok receives user reports of violative Hate Speech, it is actioned efficiently. Under 'Hate Speech & Hateful Behaviour', 144,394 of the 184,660 total videos removed following user reports were removed within two hours of receiving the report. Collectively, TikTok's proactive detection and reporting tools have worked effectively to ensure that 864,390 of the 1,083,469 total videos removed under 'Hate Speech & Hateful Behaviour' were removed within 24 hours of upload.

The effectiveness, reasonableness and proportionality of TikTok's moderation of Hate Speech is indicated by its appeals data, as users can submit appeals against content removals if they believe TikTok has made a mistake. Under the 'Hate Speech & Hateful Behaviour' policy, only 136,526 of the 1,083,469 total videos removed were successfully appealed and reversed.

4. Key stakeholder engagement:

For detailed information on how TikTok has engaged with external stakeholders to inform its approach to risk identification and risk mitigation, please refer to Annex 4 This Annex provides an in-depth overview of the collaborative efforts and consultations that have shaped TikTok's strategies in these areas.

5. Prioritisation:

When considered in the context of world events, the prevention of the dissemination of Hate Speech content and the mitigation of the evolving risks to users and society remain a key priority for TikTok. TikTok has reported above on its progress in the last year, and below it states key additional actions for the year ahead. TikTok has closely considered its risk environment and the inherent and residual risk discussed above. TikTok plans to devote extra resources to combating Hate Speech in the year ahead and as a result it remains a Tier 2 risk. As noted in Year 1, TikTok will continue to keep its prioritisation under review and remains a committed signatory to the revised Hate Speech Code.

6. Planned further mitigation effectiveness improvements:

- **Article 35(1)(b), Policy review:** TikTok plans to execute a full Policy review to relaunch its Hateful Behaviour Policies to support all moderation teams. This will include the following policy titles: Hate Speech & Hateful Ideologies, Slurs, Negative Stereotypes and Generalisations, and Disparaging Religion. This project aims to increase moderation accuracy and reduce average handling time for moderation of Hate Speech content;
- **Article 35 (1)(c), Extension of specialised human moderation processes:** TikTok plans to extend its hateful speech specialised moderation process to more EU markets in order to handle nuances of hateful speech content. These will launch in Spain, Poland, France, Netherlands and Sweden;

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014614

- **Article 35 (1)(c), Explore better model usage on content detection**: TikTok proposes to explore how to use models to capture nuanced and contextual hate globally, including EU countries; and
- **Article 35 (1)(i), Roll out additional features to empower users against hate:** TikTok will develop and launch additional features including a 'Creator Care Mode' to empower users to protect themselves against hateful behaviour. It will allow creators to filter unfriendly comments, comments from repeat violators and filter recommended keywords.

7. ILLEGAL CONTENT: RISKS OF TERRORIST CONTENT

1. Description of the risk:

a. Description of the risk from Year 1

- TikTok understands and interprets the term **'Terrorist Content'** in a manner consistent with EU and EU member state law, and its scope has been defined, in particular, having regard to Regulation (EU) 2021/784 on addressing the dissemination of Terrorist Content Online (**'TCO Regulation'**), and Directive 2017/541 (EU), Articles 3 to 12. Those laws outline the following illegal activities: (i) Terrorist offences; (ii) Offences relating to a terrorist group; (iii) Public provocation to commit a terrorist offence; (iv) Recruitment for terrorism; (v) Providing training for terrorism; (vi) Receiving training for terrorism; (vii) Travelling for the purpose of terrorism; (viii) Organising or otherwise facilitating travelling for the purpose of terrorism; (ix) Terrorist financing; and (x) Other offences related to terrorist activities;
- The risk may arise from users attempting to share or disseminate the following content on or through the Platform including through video, livestream, comments and in profile information:
    - Content that praises, promotes, glorifies, or supports violent acts or extremist organisations or individuals, or content seeking to raise funds or obtain assistance for such entities;
    - Content that encourages participation in, or intends to recruit individuals to, violent extremist organisations; and/or
    - Content with names, symbols, logos, flags, slogans, uniforms, gestures, salutes, illustrations, portraits, songs, music, lyrics, or other objects meant to represent violent extremist organisations or individuals.
- TikTok also appreciates that Terrorist Content is in direct violation of the right to life, liberty and security Art. 2 of the Charter which enshrines the right to life, and that Art. 6 protects the right to liberty and security, and that such rights may be undermined by the dissemination of Terrorist Content on the Platform. TikTok also recognises that efforts to moderate Terrorist Content must be accurate, balanced, and reasonable to ensure that such efforts do not disproportionately impact on other fundamental rights under the Charter, in particular the rights to freedom of expression and information, data protection and non-discrimination, and freedom of thought, conscience and religion, as well as TikTok's freedom to conduct a business.

b. Risk Identification: changes to the risk profile since Year 1

In its Year 1 Risk Assessment, TikTok classified Terrorist Content as a **Tier 2 priority**.

This section summarises additional risks identified in Year 2. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk detection processes, and extensive consultations with external stakeholders.

Summary of risks identified in Year 1

TikTok identified the following inherent risks as part of its Year 1 Risk Assessment:

- **Exposure to illegal, violent, and offensive content:** The risk that users' exposure to harmful content can have negative effects on mental health and potentially lead to desensitisation to violence. Some users that are susceptible to mental illness may also be vulnerable to radicalisation pathways;
- **Content moderation evasion tactics**: The risk that terrorist and violent extremist entities routinely evolve the way they communicate online, including through otherwise benign symbols and vocabulary. There is a risk that TikTok's content moderation systems may not be familiar with vocabulary related to Terrorist Content where it is rapidly evolving; and
- **Intentional manipulation of TikTok via impersonation:** The risk that users can manipulate their account to pose as another individual in a deceptive manner. Impersonation may enable terrorist organisations to create false accounts that appear trustworthy which can result in the recruitment and promotion of radical or extremist beliefs.

Additional risks identified in Year 2

In Year 2, TikTok identified the following additional risk areas, which it considers to be reflective of the dynamic and evolving nature of Terrorist Content:

- **Risks related to the evolving nature of terrorist use of UGC platforms***:* TikTok continues to monitor and evaluate the risks identified in its Year 1 Risk Assessment. In Year 2, TikTok identified several priority areas reflective of the dynamic and evolving nature of how terrorist and violent extremist entities innovate to exploit UGC-platforms. Notably, the Hamas terror attacks of 7 October 2023 and the resulting Israel-Hamas war were significant events that shaped these priorities. As of April 2024, TikTok has removed more than 3.1 million videos and suspended over 140,000 livestreams in Israel and Palestine for violating Community Guidelines, including content related to promoting Hamas, hate speech, violent extremism, and misinformation. Over the past year, TikTok saw a rise in prevalence of Osama Bin Laden's manifesto, 'Letter to America' (the '**Letter**'). Elements of the manifesto included the dissemination of violent extremist content, discussing the conflict in a manner that initially evaded content moderation. This has highlighted the risk of not all violent extremist manifestos being automatically detected and removed, particularly when parts of the manifesto are used to discuss non-violative topics. The promotion and discussion of the Letter on TikTok was problematic, as it could be mis-used to further extremist agendas, promote violence, or incite fear and hatred without necessary context and condemnation. Discussions on sensitive issues related to Terrorist Content must be handled

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014616

responsibly, and the uploader must provide counter-speech, condemnation, or awareness of the harms of terrorism;

- **Evolving use of multiple platform features to share violative content**: TikTok has identified a growing trend of users exploiting account biographies and profile information to disseminate content that violates its Community Guidelines on violent extremism. Violative behaviour among violent extremist users is not limited to a single type of content, but is spread across various platform features. Notably, these users frequently use videos, comments, and user profiles to propagate their messages. This issue has been highlighted through feedback from experts, escalations concerning terrorist and violent extremist activities, and an internal analysis of policy breaches; and
- **Advertising policies**: Prior to the Israel-Hamas war, TikTok experienced a low incidence of content in ads related to terrorism or terrorist organisations. Although this remains uncommon, there has been an increase in ads with humanitarian messages related to the conflict. TikTok has recognised a potential concern regarding ads that could reference terrorism or related activities, potentially seeking to exploit the situation. Moreover, there has been a rise in ads requesting donations, underscoring the need for thorough verification processes. While neither ads with humanitarian messages nor ads requesting donations are inherently risky, TikTok's existing advertising policies were nevertheless reassessed to ensure the platform's policies adequately address these complexities without suppressing calls for humanitarian aid and support.

c. Inherent Risk in Year 2

For a detailed analysis on how TikTok assessed the baseline severity and probability for Terrorist Content in Year 1, please refer to TikTok's Year 1 Systemic Risk Assessment Report.

In Year 2, TikTok assesses the overall severity of Terrorist Content to be 'Material' and the overall probability to be 'Possible' in Year 2. Taking this into account, TikTok assesses the inherent risk (that is, the risk without any mitigations in place) (that is, the risk without any mitigations in place) for Terrorist Content to be 'Moderate'.

2. Mitigation Measures:

For detailed information on TikTok's system of mitigations/controls (under DSA Art. 35 (1)(a)-(k)) that applied when its designation as a VLOP came into force, please refer to the Year 1 Risk Assessment Report. That Year 1 Report identified further mitigation improvements, progress on which (and any other such mitigations) is listed below. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk monitoring and detection processes, and extensive consultations with external stakeholders.

Implementation of Additional Mitigation Effectiveness Measures in Year 2	
MEASURES IN ACCORDANCE WITH DSA ART. 35(1) A-K	**DESCRIPTION**

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

(b) Adaptation of terms and conditions	TikTok updated its Community Guidelines to include the 'sharing of manifestos' of violent extremist entities as an important example of content not allowed on the Platform. TikTok engaged its Safety Advisory Councils across the globe, as well as experts in antisemitism and Islamophobia to ensure Community Guidelines were appropriate within the context of the outbreak of violence to ensure policies and corresponding enforcement were and remain proportionate to the threat. TikTok also convened special meetings with priority partners ranging from international organisations such as Tech Against Terrorism and the International Committee of the Red Cross to identify violative content, as well as local and regional NGOs with knowledge of cultural trends and nuances.
(c) Adaptation of content moderation processes	In addition to the points in (a) above and as a further response to the events of 7 October 2023, TikTok deployed 72 additional Arabic and Hebrew speaking moderators in order to augment existing content moderation teams in reviewing content, checking for new keywords[9] and assisting with translations in relation to the Israel-Hamas war. TikTok is in the process of globally scaling specialised processes dedicated to handling Terrorist Content. TikTok has undertaken a trial of its revised process and training and has determined that they have operated to increase enforcement quality with no identified negative impact on free speech. In the long term, these processes will support the strengthening of institutional subject matter expertise and knowledge and provide higher quality training data for machine learning models for automatic moderation. As described above, TikTok's Community Guidelines were updated to include the dissemination of manifestos from hateful organisations as an example of prohibited content. TikTok implemented this update to enable it to remove depictions or amplification of such manifestos where the content lacks contextual relevance or are shared outside of a legitimate public interest context. TikTok promptly implemented measures to address the 'Letter to America', including the following actions: • TikTok implemented proactive detection for the text of the manifesto (in whole or in part);

[9] Keywords are sensitive words, written as a single lexical word or a text string written with or without a blank between adjacent words. It can also consist of emoji/s or special character/s. The keywords trigger enforcement action. The purpose is to trigger manual review or automatically reject a term across numerous text-based features including but not limited to block search, delete or downrank comments, block username creation, prevent audio creation.

	• TikTok prevented users from searching relevant keywords, including suggested searches for the Letter, to prevent discoverability of any iterations of the content; • TikTok used detection tools to identify and remove content similar to previously flagged violative material containing the Letter; • TikTok set up a process to automatically delete videos that matched known harmful content, leading to 660 deleted pieces of content, each of which was deleted with no views by users; • TikTok prohibited the use of related keywords as hashtags; and • TikTok prevented the audio read-outs of the Letter to be included in any other content. TikTok continues to monitor the conflict closely and will continue to adapt moderator guidance and enforcement processes as needed to address evolving issues. TikTok has also updated its moderator guidance across violent extremism, hateful behaviour, illegal activities and regulated goods multiple times since the start of the conflict to ensure TikTok moderators are as well-equipped as possible to identify and action violating content related to the conflict and that enforcement reflects TikTok's Community Guidelines. TikTok reviewed existing moderator guidance on violent and hateful organisations and individuals to ensure entities violating Community Guidelines are promptly and thoroughly removed from the Platform. TikTok worked closely with external partners to update moderator guidance to ensure alignment with enforcement approaches. TikTok also convened the first issue-specific Safety Advisory Council meeting in December 2023 to focus on responses to the Israel-Hamas war.
(d) Adaptation of algorithmic systems	TikTok has been working on improving its system that automatically removes content that violates TikTok's violent extremism policies, or directs it for human review, if the content appears violating but is below a certain confidence threshold. This allows TikTok to reduce the reach of violating content. For example, data on TikTok's enforcement of its Community Guidelines in the EU shows that TikTok removed 247,114 videos without any views under its 'Violent & Hateful Orgs & Individuals' policy, and 61,194 videos without any views under its 'Violent Behaviours & Criminal Activities' policy between Q3 2023 and Q2 2024.
(e) Adaptation of advertising systems	On 16 October 2023, TikTok updated its advertising policies, permitting only recognised NGOs to advertise fundraising initiatives related to the Israel-Hamas war. These NGOs must verify their identity before they are able to place ads on TikTok. Ads must not reference

	terrorist organisations or violate any other existing Community Guidelines or advertising policies. In February 2024, TikTok updated its advertising policies to allow ads to advocate for victims related to the Israel-Hamas war, provided that these ads do not violate other advertising policies in doing so.
(f) Reinforcing risk detection measures	As mentioned in Year 1 as a key mitigation improvement, TikTok continues to monitor risks associated with Terrorist Content and improve its detection capabilities. In early 2024, TikTok updated its existing risk management processes to better identify trends in emerging risks to the platform. This process now systematically identifies, categorises, and prioritises online trends on a recurring basis. It also establishes oversight and accountability for mitigation efforts aimed at containing these trends and risks. This improved process ensures that both short-term strategies (such as escalation management) and long-term solutions (such as detection iteration) work together effectively. Since the revamp, TikTok has processed 57 high-risk trends, ensuring the removal of associated harmful content, making it undiscoverable through search, providing updated guidance to moderation teams, and updating proactive detection mechanisms. When a new trend under Terrorist Content policies is identified, relevant stakeholders are brought together to discuss mitigation strategies, understand the root cause of the content appearing, and implement short-term plans to control the trend and minimise user exposure.
(g) Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21	TikTok remains ready to engage with EU 'trusted flaggers' as they continue to be designated (read section 5) and in the meantime, TikTok continues to receive reports from Community Partners and subject matter experts.

Case study: Israel-Hamas war

TikTok activated its crisis protocols immediately in response to the Hamas attack on 7 October 2023, swiftly establishing 24/7 control rooms and mobilising over 40,000 safety professionals worldwide. It

was clear from the start that content about the conflict would be visible to, and affect, users in Europe. Additional moderation resources in Arabic and Hebrew were deployed swiftly to prevent the spread of violent, hateful, or misleading content. TikTok's Crisis Management playbook, consisting of six phases of detection, triaging and readiness, activation, response, de-escalation and recovery, and closure guided the response. By 6:30 am (Irish time) on Saturday 7 October 2023, the detection phase was activated, and within minutes, the event was escalated for assessment.

Application of TikTok's Crisis Protocol to the Israel-Hamas war

By 7:25 am on 7 October, TikTok declared the situation a crisis and moved to the activation phase. By 7:30 am, a cross-functional response team was deployed, initiating strategies such as 24/7 moderation coverage in Arabic and Hebrew. This included creating dedicated groups to manage the crisis and additional reviews of uploads for violative content. A Crisis Committee, including senior leadership from the EMEA region, was established by 10:00 am, and senior global leadership was informed due to the conflict's severity and global impact.

TikTok's response was driven by its command centre, consisting of 200 members from 14 global teams, primarily operating out of Dublin, with members also located in Israel, the Middle East, and the USA. The command centre ensures Platform safety, coordinates crisis management across Trust & Safety teams, provides regular updates to senior leadership, and identifies and implements both short-term and long-term risk mitigation strategies. The team remains vigilant during any declared crisis, addressing fluctuations in content moderation and emerging risks to maintain the Platform's integrity during the ongoing Israel-Hamas war.

On 9 October 2023, based on the evolution of the crisis, TikTok decided to implement restrictions on content by suspending user-generated short-form video and LIVE content originating from Israel and Palestine from being eligible for recommendation on the For You Page for users in Europe. This was a preventative measure to mitigate the risk of violative LIVE and other video content (e.g. Hate Speech, Terrorist Content) being widely disseminated on the Platform. TikTok also introduced new restrictions on LIVE eligibility for users based in Israel and Palestine, to mitigate the risk of bad actors using the Platform to post violative content.

The de-escalation and recovery phase of any crisis involves daily reviews of information sources and emerging trends to monitor the crisis's evolution and TikTok's response. As the volume of violative content stabilises, the frequency of meetings and updates to senior leadership is reduced, but the cross-functional team continues monitoring for new risks. In the closure phase, a root cause analysis is conducted once the crisis impact on the Platform diminishes and metrics return to pre-crisis levels.

3. Residual Risk in Year 2:

Following an assessment of the effectiveness, reasonableness and proportionality of TikTok's mitigations relevant to the systemic risk of Terrorist Content, TikTok has assessed residual risk to be 'Moderate' in Year 2. TikTok commits to expand and invest in mitigation measures to identify and mitigate Terrorist Content on the Platform.

This assessment was conducted on the basis of qualitative and quantitative insights, including data on TikTok's enforcement of its Community Guidelines in the EU for the period of Q3 2023 to Q2 2024.

Under the Community Guidelines policy of 'Violent & Hateful Orgs', TikTok removed 369,712 total videos in the EU, with 346,449 detected and removed proactively, and 247,114 removed without any views. Under the 'Violent Behaviours & Criminal Activities' policy, TikTok removed 163,006 total videos in the EU, with 138,569 detected and removed proactively, and 61,194 removed without any views. This high proportion of proactive removal suggests that TikTok's content moderation systems remain effective in detecting and removing Terrorist Content on the Platform, helping to mitigate the systemic risk.

This assessment of the effectiveness of TikTok's proactive detection systems is further informed by the following data on user reports, which serve as a useful indication of the volume of Terrorist Content that is not proactively detected and removed by TikTok. User reports accounted for a minority of Terrorist Content detected and removed by TikTok, at 23,263 of the 369,712 total videos detected and removed under 'Violent & Hateful Orgs & Individuals', and 24,437 of the 163,006 total videos removed under 'Violent Behaviours & Criminal Activities'. TikTok remains committed to continuously working to improve and iterate its proactive detection capabilities, while noting that Terrorist Content is a rapidly evolving risk area, with nuance across regions and languages.

TikTok's user reporting tools are an important component of its content moderation process, ensuring that Terrorist Content is detected and removed, including new and evolving forms of Terrorist Content. The data demonstrates that when TikTok receives user reports of violative Terrorist Content, it is actioned efficiently. Under 'Violent & Hateful Orgs & Individuals', 18,384 of the 23,263 total videos removed following user reports were removed within two hours of receiving the report. Similarly, under 'Violent Behaviours & Criminal Activities', 20,211 of the 24,437 total videos reported were removed within two hours. Collectively, TikTok's proactive detection and reporting tools have worked effectively to ensure that 300,163 of the 369,712 total videos removed under 'Violent & Hateful Orgs & Individuals' were removed within 24 hours of upload. Similarly, 123,204 of the 163,006 total videos removed under 'Violent Behaviours & Criminal Activities' were removed within 24 hours.

The effectiveness, reasonableness and proportionality of TikTok's moderation of Terrorist Content is indicated by its appeals data, as users can submit appeals against content removals if they believe TikTok has made a mistake. Under the 'Violent & Hateful Orgs & Individuals' policy, only 45,117 of the 369,712 total videos removed were successfully appealed and reversed. Similarly, under 'Violent Behaviours & Criminal Activities', only 17,450 of the 163,006 total videos removed were successfully appealed and reversed.

4. Key stakeholder engagement:

For detailed information on how TikTok has engaged with external stakeholders to inform its approach to risk identification and risk mitigation, please refer to Annex 4. This Annex provides an in-depth overview of the collaborative efforts and consultations that have shaped TikTok's strategies in these areas.

5. Prioritisation:

TikTok has reported above on its progress in the last year, and below it states key additional actions for the year ahead. TikTok has closely considered its risk environment and the inherent and residual risk discussed above. TikTok plans to devote extra resources to combatting the risk of Terrorist Content in

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

the year ahead and as a result it remains a Tier 2 risk. As noted in Year 1, TikTok will continue to keep its prioritisation under review.

6. Planned further mitigation effectiveness improvements:

- **Article 35(1)(b), Enhanced policies:** TikTok plans to review and enhance four of its policies to better address terrorist and violent extremist content. This will include updating the policies around how we identify dangerous entities, dangerous political entities, credible threats of severe violence and violent behaviours.
- **Article 35(1)(c), Specialised moderation:** TikTok plans to extend specialised moderation processes for Terrorist Content to more EU markets to ensure nuances are appropriately addressed. TikTok will expand to Spain, Poland, France, Netherlands, and Sweden.
- **Article 35(1)(c), Content moderation:** TikTok will build an end-to-end enforcement process to effectively remove newly identified terrorist organisations and individuals.
- **Article 35(1)(e), Adapting advertising systems:** TikTok will continue to ensure the readiness of its human and automated moderation systems to prevent emerging and evolving Terrorist Content from being monetized on the Platform.

8. ILLEGAL CONTENT: RISKS OF INTELLECTUAL PROPERTY INFRINGING CONTENT

1. Description of the risk:

a. Description of the risk from Year 1

- TikTok understands the term **'IP-Infringing Content'** to mean content that is created and disseminated in breach of copyright or other intellectual property (**'IP'**) rights;
- TikTok's approach to actioning IP-Infringing Content is structured in particular in accordance with relevant EU laws on copyright, including Directive (EU) 2019/790 (the **'Copyright Directive"**);
- IP-Infringing Content includes the following content, whether uploaded as a video, livestream, or in profile information on the Platform:
    - Content that reproduces, and disseminates on the Platform, the original work of another person or entity without that person's or entity's permission and which does not fall in one of the copyright exceptions. This content may include music works and audio files, artistic works (e.g. photographs, paintings, drawings, and other original visual renderings), and audio-visual recordings;
    - Content that contains the unauthorised use of a trademark or service mark in connection with goods or services in a way that is likely to cause confusion, deception or mistake about the source, origin, sponsorship or affiliation of the associated goods and/or services; or
    - Content that advertises or promotes counterfeit products.
- In addition, TikTok also notes that Art. 17 of the Charter enshrines the right to protection of intellectual property.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014623

b. Risk Identification: changes to the risk profile since Year 1

In its Year 1 Risk Assessment, TikTok classified IP-Infringing Content as a **Tier 3 priority** in its priority tiering system.

This section summarises additional risks identified in Year 2. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk detection processes, and extensive consultations with external stakeholders.

Summary of risks identified in Year 1

TikTok identified the following inherent risks as part of its Year 1 Risk Assessment:

- **Sharing and dissemination of IP-Infringing Content**: The risk that users disseminate IP-Infringing Content;
- **Terms of Service and Community Guidelines:** The risk that these rules may be misinterpreted by users (intentionally or unintentionally), leading to a risk of users uploading and disseminating IP-Infringing Content on the Platform;
- **Moderation of IP-Infringing Content:** The risk that TikTok may inaccurately remove content that the rightsholder has informed TikTok that it in fact wishes to remain online; and
- **Ads**: The risk that advertisers share ads content which disseminates IP-Infringing Content.

Additional risks identified in Year 2

In Year 2, TikTok identified the following additional risk areas, which it considers to be reflective of the dynamic and evolving nature of IP-Infringing Content.

- **AIGC***:* As third party AIGC technology becomes more accessible and sophisticated, TikTok has observed an increase in claims that AIGC posted by users is IP-Infringing Content. There are challenges with moderating this type of content, for example, it can be difficult for rights holders to demonstrate that their copyright has been infringed by AIGC content, and therefore difficult for moderators to make an assessment on infringement; and
- **Counterfeit goods***:* The risk that counterfeit goods are marketed to users on the Platform, in violation of the Community Guidelines. Bad actors may provide instructions to users on TikTok directing them to third-party platforms where counterfeit goods can be purchased. Counterfeit goods may pose potential health and safety risks to consumers.

c. Inherent Risk in Year 2

For a detailed analysis on how TikTok assessed the baseline severity and probability for IP-Infringing Content in Year 1, please refer to TikTok's Year 1 Systemic Risk Assessment Report.

In Year 2, TikTok assesses the overall severity of IP-Infringing Content has been assessed to be 'Moderate'. TikTok assesses the probability of IP-Infringing Content in Year 2 to be 'Possible'. Taking into account the severity and probability in Year 2, TikTok has assessed the inherent risk (that is, the risk without any mitigations in place) of IP-Infringing Content being shared on the Platform in Year 2 to be 'Moderate'.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014624

2. Mitigation measures:

For detailed information on TikTok's system of mitigations/controls (under DSA Art. 35 (1)(a)-(k)) that applied when its designation as a VLOP came into force, please refer to the Year 1 Risk Assessment Report. That Year 1 Report identified further mitigation improvements, progress on which (and any other such mitigations) is listed below. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk monitoring and detection processes, and extensive consultations with external stakeholders.

Implementation of Additional Mitigation Effectiveness Measures in Year 2	
MEASURES IN ACCORDANCE WITH DSA ART. 35(1) A-K	**DESCRIPTION**
(a) Adaptation of feature or platform design	TikTok has made changes to the way IP-Infringing Content can be manually discovered on the Platform by blocking specific search terms. When users search for certain terms, no results will be shown, and they will receive a message notifying them that unauthorised streaming of digital content and promotion of counterfeit goods violates TikTok's Community Guidelines. To help mitigate the risk of bad actors marketing counterfeit goods on the Platform, TikTok has prevented users from adding website links to their profiles that lead to known sites which sell or promote counterfeit content or facilitate digital piracy (i.e. illegal streaming/torrent websites). TikTok has optimised its external webforms and internal systems to streamline the process through which a rightsholder or reporter can request that future copies of reported and removed infringing videos are prevented from reappearing on the Platform. TikTok's copyright webform has also been improved to reduce the incidence of rights holders or reporters providing insufficient evidence.
(b) Adaptation of terms and conditions	TikTok has launched a dedicated Music Terms of Service to help users understand how music can be legally used on the Platform.
(c) Adaptation of content moderation processes	TikTok has adapted its content moderation workflows for IP notice and takedown requests to mitigate potential risks associated with AIGC. This has included: • Establishing an internal guideline for processing IP reports against AIGC music; • Revising standard operating procedures to instruct moderators on how to handle a piece of AIGC reported to TikTok for copyright or trademark infringement;

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014625

	• Guidance has been provided to moderators to help them better identify when reports of IP-Infringement have been made against AIGC; and • Moderators have been advised on scenarios involving claims against AIGC that require escalation to legal teams for further risk assessment. This includes grey area 'edge cases' that may be difficult for a moderator to assess without specialised legal knowledge. TikTok will remain vigilant as intellectual property law evolves in response to AIGC and will implement updates to its policies where appropriate.
(e) Adaptation of advertising systems	TikTok has improved its automated counterfeit ad detection system so that it can now identify brands in ads and notify moderators of their actual lowest prices. This supports human moderators to assess whether ads are counterfeit, based on unreasonably low prices.
(f) Reinforcing risk detection measures	As mentioned in the Year 1 Report as a key planned mitigations effectiveness improvement, TikTok has continued to further develop its internal tools to detect and prevent copyright infringing activities. These improvements include: • Based on the trends and characteristics of counterfeit goods sales ads, TikTok has provided regular training to moderators, alongside regular improvements to the Platforms automated detection tools; • TikTok has optimised the complaint path for IP rights holders, making it clearer and more accessible, and iterated the complaint interface on its web page to make it easier to understand. TikTok has also fixed some known bug issues to make the complaint-filing process smoother; and • TikTok has optimised its internal ticketing system for receiving and responding to IP-infringement reports, improving its efficiency by revising policies and workflow, and simplifying documentation. Additionally, TikTok has provided more rightsholders with access to TikTok's copyright tools, which ensure the unavailability of protected works on the Platform where rightsholders have provided TikTok with the relevant and necessary information. Blocked content cannot be accessed by users in the region(s) in which it is blocked.
(g) Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the	TikTok remains ready to engage with EU 'trusted flaggers' as they continue to be designated (read section 5). TikTok treats reports from the Copyright Information and Anti-Piracy Centre (**'CIAPC'**) with priority since February 2024. This was in advance of the Finnish non-profit appointed as a Trusted Flagger by the Finnish Digital Services Coordinator on 7

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014626

decisions of out-of-court dispute settlement bodies pursuant to Article 21	March 2024. To date CIAPC has not submitted any reports under Article 16 DSA. For detailed information on how TikTok has engaged with external stakeholders to inform its approach to identifying and mitigating the risk of Intellectual Property Infringing Content, please refer to Annex 4.

3. Residual Risk in Year 2:

Following an assessment of the effectiveness, reasonableness and proportionality of TikTok's mitigations relevant to the systemic risk of IP-Infringing Content, TikTok has assessed residual risk to be 'Low' in Year 2. TikTok commits to continuing to expand and invest in mitigation measures to identify and mitigate IP-infringing Content on the Platform.

This assessment was conducted on the basis of qualitative and quantitative insights, including global data from TikTok's Intellectual Property Removal Requests Report[10] for the period of 1 July to 31 December 2023. TikTok provides reporting tools to rights holders and their representatives through which they can notify TikTok of potential copyright or trademark infringements and, where appropriate, have the content taken down. In the reporting period detailed above, TikTok received 272,555 total copyright removal requests globally, with 56.1% assessed as successful. In the same period, TikTok received 29,188 trademark removal requests globally, with 57.4% assessed as successful.

This data indicates the volume of copyright or trademark violating content that is not detected and removed by TikTok's proactive systems. At 152,830 total successful copyright removal requests and 16,758 total successful trademark removal requests for the period 1 July to 31 December 2023, this represents an extremely small share of overall content uploaded and shared on the Platform. TikTok is committed to proactively preventing IP-infringing content from being uploaded to the Platform, and efficiently and diligently processing removal requests from appropriate parties.

4. Key stakeholder engagement:

For detailed information on how TikTok has engaged with external stakeholders to inform its approach to risk identification and risk mitigation, please refer to Annex 4. This Annex provides an in-depth overview of the collaborative efforts and consultations that have shaped TikTok's strategies in these areas.

5. Prioritisation:

TikTok has reported above on its progress in the last year, and below it states key additional actions for the year ahead. TikTok has closely considered its risk environment and the inherent and residual risk discussed above. TikTok plans to devote extra resources to combating IP-Infringing Content in the year ahead and as a result it remains a Tier 3 risk. As noted in Year 1, TikTok will continue to keep its prioritisation under review.

[10] https://www.tiktok.com/transparency/en/intellectual-property-removal-requests-2023-2/.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014627

6. Planned further mitigation effectiveness improvements:

- **Article 35(1)(i), Awareness raising measures:** TikTok is committed to improving its approach to mitigate IP-Infringing Content by creating additional resources to educate users who receive infringement notices and enhancing collaboration with industry groups to improve the reporting process for rights-holders;
- **Article 35(1)(c), Reviews of Guidelines:** In addition to current regular policy reviews, TikTok will implement a regular review for Community Guidelines updates and moderator guidance to ensure effectiveness and relevance;
- **Article 35(2)(e), Adapting advertising systems:** TikTok is planning to implement new stringent measures to identify and remove users who consistently violate the Platform's ad policies in respect of IP, including implementing enforcement actions against repeat offenders and fraudulent behaviours. TikTok also commits to regularly updating its internal ad policies and processes to reflect changes in industry standards and legal requirements regarding IP. Last, TikTok is planning to develop and maintain a process for expeditiously receiving and addressing rights holders' complaints regarding IP-infringing content in ads; and
- **Article 35 (1) (f), Processing illegal content orders:** Following development of a robust Intellectual Property system for high risk events scenario-planning and readiness, TikTok will continue to strengthen communication channels with rightsholders and cross-functional working groups within Trust and Safety.

9. ILLEGAL CONTENT: RISKS OF CHILD SEXUAL ABUSE MATERIAL AND CHILD SEXUAL EXPLOITATION AND ABUSE

1. Description of the risk:

a. Description of the risk from Year 1

- TikTok understands the terms child sexual abuse material ('**CSAM**'[11]) and child sexual exploitation and abuse ('**CSEA**') in a manner consistent with EU and EU member state laws. The scope of CSAM and CSEA risks have been defined in particular having regard to Directive 2011/93/EU, Articles 3 to 7 respectively in relation to: offences concerning sexual abuse; offences concerning sexual exploitation; offences concerning child pornography; solicitation of children for sexual purposes; and incitement, aiding and abetting, and attempts to commit such offences.
- TikTok considers that the dissemination of CSAM content may involve users of the Platform attempting to, whether in relation to video, photo or livestream content or in relation to the abuse of account settings:
    - share, re-share or offer to trade or sell, or direct users of the Platform to obtain or distribute CSAM content;

[11] Although various legal texts still refer to the term 'child pornography', in line with the Guidelines for the Protection of Children from Sexual Exploitation and Sexual Abuse (known as the 'Luxembourg Guidelines'), TikTok considers CSAM to be the more appropriate term.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

- generate and/or share self-generated CSAM;
- disseminate content that depicts, solicits, glorifies, or encourages child abuse imagery including nudity, sexualised minors, or sexual activity with minors;
- disseminate content that depicts, promotes, normalises, or glorifies pedophilia or the sexual assault of a minor; or
- post information on their account profile, including their username, handle, profile picture and profile bio, that contains CSAM.

- TikTok considers that the activities associated with CSEA behaviour may involve adult users of the Platform attempting to:
  - build an emotional relationship with a minor in order to gain the minor's trust for the purposes of future or ongoing sexual contact, sexual abuse, trafficking, or other exploitation;
  - solicit real-world contact between a minor and an adult or between minors with a significant age difference;
  - solicit minors to connect with an adult on another online platform, website, or other digital space for sexual purposes;
  - solicitation of nude imagery or sexual contact, through blackmail or other means of coercion; or
  - post information on their account profile, including their username, handle, profile picture and profile bio that involves CSEA behaviour.
- TikTok notes that Article 34 of the United Nations Convention on the Rights of the Child enshrines the right of the child to protection from all forms of sexual exploitation and sexual abuse, and that such risks should be assessed with the best of interests of the child as the primary consideration, in accordance with Article 24 of the Charter. TikTok notes that in 2021, the United Nations Committee on the Rights of the Child underlined that these rights must be equally protected in the digital environment.[12]

b. Risk Identification: changes to the risk profile since Year 1

In its Year 1 Risk Assessment, TikTok classified CSAM and CSEA as a **Tier 1 priority** in its priority tiering system.

This section summarises additional risks identified in Year 2. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk detection processes, and extensive consultations with external stakeholders.

Summary of risks identified in Year 1

TikTok identified the following risks related to CSAM and CSEA in its Year 1 Risk Assessment:

- Sharing and dissemination of CSAM: The risk of users attempting to share and disseminate CSAM on the Platform; and
- Carrying out or participating in CSEA: The risk of users attempting to use the Platform to carry-out or participate in CSEA.

[12] UN General Comment No. 25 (2021) on Children's Rights in Relation to the Digital Environment.

Additional risks identified in Year 2

In Year 2, TikTok identified the following additional risk areas, which it considers to be reflective of the dynamic and evolving nature of CSAM and CSEA.

- **Sextortion and grooming**: Sextortion is a growing threat to Younger Users of online platforms across the globe, with data demonstrating a rise in abuse targeting teenage boys and young adult males[13]. Grooming also continues to be a risk to Younger Users of all genders. Sextortion and grooming can lead to online and offline risks to the safety and well-being of users and encourage the proliferation of CSAM production and distribution networks. The incidence of sextortion and grooming on TikTok is low, but these risks may be challenging to detect and moderate given the following factors:
    - **Subtle tactics:** Groomers build trust over time, making their conversation (in comments related to UGC) hard to distinguish from regular interactions. Automated systems often struggle with the nuanced and contextual nature of these conversations;
    - **Anonymity:** Predators frequently use fake profiles and multiple accounts to evade detection, and often employ mass communication tactics to reach more victims;
    - **Underreporting:** Victims may not report grooming or sextortion due to fear, shame, or lack of awareness of their ability to make such a report; and/or
    - **Abuse patterns:** TikTok has identified common abuse patterns, such as sharing off-platform CSAM links and creating 'collector accounts' that may indicate a sexual interest in children but where individual pieces of content do not violate our policies. These require holistic account assessments for effective detection.
- **Difficulty in detecting covert and suspicious behavioural signals**: Online predators may conceal their behaviour so as to bypass moderation efforts by interacting with and searching for content created by Younger Users without explicitly violating Community Guidelines. Online platforms could be exploited by adults with predatory intent, by connecting with like-minded adults via comments, likes, and follower lists, potentially forming 'offender communities'. TikTok's current systems have historically been targeted at content-level review, but their systems are now being enhanced and complemented to identify holistic account-level signals;
- **Increased volume of reporting**: As TikTok's detection systems can now identify more content that needs to be reviewed for potential reporting to the National Center for Missing and Exploited Children ('**NCMEC**'). In 2023, TikTok sent 590,376 reports to NCMEC, whereas in 2022 TikTok sent 288,125. This increase in flagged content requires a more focused approach, prioritising the most serious cases over a simple 'first in, first out' method, to ensure the swift handling of the most urgent matters. The increase in the volume of sensitive content also requires attention to ensure continued well-being of moderators through strong support systems. Central to TikTok's approach is accurate reporting and effective cooperation with law-enforcement, including requiring all NCMEC-related content to be reviewed by human moderators. This adds significant pressure on human staff and limits the use of automated processes. Although TikTok's enhanced detection has resulted in an increased volume of reporting, it remains low compared to other platforms;[14]
- **Ads containing child abuse cases:** Advertising systems may be mis-used to share ads featuring AIGC depictions of minors that were involved in real-life abuse or abduction cases.

[13] Source: Missingkids.org

[14] https://www.missingkids.org/content/dam/missingkids/pdfs/2023-reports-by-esp.pdf

While these ads are removed under existing Youth Safety policies, this content is still treated as an emerging risk in Year 2.

c. Inherent Risk in Year 2

For a detailed analysis on how TikTok assessed the baseline severity and probability for CSAM and CSEA in Year 1, please refer to TikTok's Year 1 Systemic Risk Assessment Report.

In Year 2, TikTok assesses the overall severity of CSAM and CSEA to be 'Material' and the overall probability to be 'Possible' in Year 2. Taking this into account, TikTok assesses the inherent risk (that is, the risk without any mitigations in place) for CSAM and CSEA to be 'Moderate'.

Mitigation Measures:

Please refer to the Year 1 Risk Assessment Report for detailed information on TikTok's system of mitigations/controls (under DSA Art. 35 (1)(a)-(k)) that applied when its designation as a VLOP came into force. That Year 1 Report identified further mitigation improvements, progress on which (and any other such mitigations) is listed below. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk monitoring and detection processes, and extensive consultations with external stakeholders.

Implementation of Additional Mitigation Effectiveness Measures in Year 2	
MEASURES IN ACCORDANCE WITH DSA ART. 35(1) A-K	**DESCRIPTION**
(a) Adaptation of feature or platform design	In addition to blocking searches associated with CSEA and CSAM, TikTok deploys in-app warnings and deterrence resources designed by CSEA prevention organisations[15] to deter users from violative behaviour related to CSEA and CSAM.
(b) Adaptation of terms and conditions	TikTok's Community Guidelines have been updated to classify hyper-realistic AIGC sexual content which depicts minors as CSAM, ensuring its prompt removal and reporting of such content to NCMEC.
(c) Adaptation of content moderation processes	In addition to the mitigation effectiveness improvements anticipated in the Year 1 Report, TikTok has updated its AI-enabled detection system to identify and flag potential CSAM and CSEA content for review. Content formats such as music and ads that are new for these detection systems have now been integrated. TikTok also uses several third-party CSAM detection systems and has expanded its scope of auto enforcement to reduce moderation exposure to explicit CSEA content. To guard against moderation errors and moderator well-being, high precision CSAM matches (e.g. Photo DNA hashing with high accuracy matching) are auto-removed and reported to NCMEC in batch.

[15] https://www.stopitnow.org.uk/

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014631

(e) Adaptation of advertising systems	TikTok has made a number of updates to the policies and enforcement actions that govern its ads, including: • **Prohibition on Child Abuse References:** Since January 2024, TikTok prohibits references to historical child abuse cases in ads and landing pages[16] on Google Play or Apple App Store, including those made using AIGC in order to protect the dignity of victims. Exceptions are made for anti-child abuse awareness campaigns and certain movies/TV shows, documentaries/trailers, provided they comply with TikTok's Minor Safety Policies; and • **Misleading AIGC Policy**: Launched in June 2024, this policy prohibits AIGC ads that depict minors, political figures, or authoritative sources. An exception is made for 'baby generator' ads, where AIGC images of minors are allowed, but the use of real photos or identities of minors is strictly prohibited. All content, including AIGC images, must adhere to TikTok's existing Minor Safety and Adult & Sexual Content policies; and • **AIGC app ads:** Ads promoting AIGC creation tools must link to Google Play or Apple App Store pages, with the verification provided by those third parties. This mitigates the risk of bad actors promoting unregulated AIGC tools.
(f) Reinforcing risk detection measures	TikTok's Centralised Child Safety Team is composed of specialists from child safety backgrounds, including former prosecutors, government, intelligence, and law enforcement officials. This team and its internal partners handle the moderation and escalation of potential CSAM content. In Year 2, the Child Safety Team worked on a new 'sandbox model'[17] that tests off-platform CSAM links without viewing content, safeguarding victims' and moderators' well-being. Trend management tools are used and aim to provide timely and efficient solutions for escalation handling with comprehensive video recall conditions and precise strategy quality and lifecycle management. As mentioned in the Year 1 Report as a key mitigation improvement, TikTok has recently updated its specialised 'Computer Vision' model, an AI-powered detection system designed to identify visual material potentially containing CSAM and CSEA. This system directs flagged content to TikTok's Child Safety Team and for review and to report to NCMEC. Content formats such as music and ads that are new for these detection systems have now been integrated, ensuring comprehensive coverage.

[16] A landing page is a standalone page created specifically for an advertising campaign, designed to direct visitors towards a clear action such as downloading an app.

[17] A sandbox model is a safe and isolated testing environment.

(g) Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21	As mentioned in the Year 1 Report, TikTok remains ready to engage with EU 'trusted flaggers' as they continue to be designated (read section 5). TikTok continues to receive reports from Community Partners and subject matter experts, including a system for NGOs and other partner entities to report CSAM and CSEA for expedited decision making. This system is managed by TikTok's Incident Management team and within the Law Enforcement channel. In the first half of 2024, 541 escalations took place from law enforcement authorities to the Child Safety Team for the EMEA (Europe, Middle East and Asia) region.
(i) Awareness raising measures	TikTok is committed to the 'Voluntary Principles to Counter Child Sexual Exploitation and Abuse' from the We Protect Global Alliance, of which it is a signatory in order to help build collaboration across platforms. TikTok's active participation in the WeProtect Global Alliance fosters collaboration across sectors – including NGOs, safety firms, and governments – facilitating the sharing of knowledge and best practices, and strengthening the collective response to CSEA. For detailed information on how TikTok has engaged with external stakeholders to inform its approach to identifying and mitigating the risk of CSAM and CSEA, please refer to Annex 4.

3. Residual Risk in Year 2:

Following its assessment of the effectiveness, reasonableness and proportionality of relevant mitigations, TikTok has assessed residual risk of CSAM and CSEA being shared on the Platform to be 'Moderate' in Year 2. TikTok commits to continuing to expand and invest in mitigation measures to identify and mitigate CSAM and CSEA on the Platform.

This assessment was conducted on the basis of qualitative and quantitative insights, including data on TikTok's enforcement of its Community Guidelines in the EU for the period of Q3 2023 to Q2 2024. Under the Community Guidelines policy of 'Safety & Civility - Youth Exploitation & Abuse', TikTok removed 1,541,762 total videos in the EU, with 1,491,028 detected and removed proactively, and 1,118,478 removed without any views. Under 'Youth Safety & Well-Being - Youth Exploitation & Abuse', TikTok removed 1,289,147 total videos in the EU, with 1,238,736 detected and removed proactively, and 866,929 removed without any views. This high proportion of proactive removal suggests that TikTok's content moderation systems remain effective in detecting and removing CSAM and CSEA on the Platform, helping to mitigate the systemic risk.

This assessment of the effectiveness of TikTok's proactive detection systems is further informed by the following data on user reports, which serve as a useful indication of the volume of CSAM and CSEA that is not proactively detected by TikTok's proactive systems. User reports accounted for a minority of CSAM and CSEA detected and removed by TikTok, at 50,734 of the 1,541,762 total videos detected and removed under 'Safety & Civility - Youth Exploitation & Abuse' and 50,411 of the 1,289,147 total videos

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014633

detected and removed under 'Youth Safety & Well-Being - Youth Exploitation & Abuse'. TikTok remains committed to continuously working to improve and iterate its proactive detection capabilities.

TikTok's user reporting tools are an important component of its content moderation process, ensuring that CSAM and CSEA is detected and removed, including new and evolving forms of CSAM and CSEA. The data demonstrates that when TikTok receives user reports of CSAM and CSEA content, it is actioned efficiently. Under 'Safety & Civility - Youth Exploitation & Abuse', 43,616 of the 50,734 total videos removed following user reports were removed within two hours of receiving the report. Under 'Youth Safety & Well-Being - Youth Exploitation & Abuse', 43,290 of the 50,411 total videos removed following user reports were removed within two hours of receiving the report. TikTok's proactive detection and reporting tools have worked effectively to ensure that 1,355,823 videos were removed under TikTok's 'Safety and Civility - Youth Exploitation & Abuse' policy and 1,103,357 videos were removed under 'Youth Safety & Wellbeing - Youth Exploitation & Abuse' within 24 hours of upload.

The effectiveness of TikTok's moderation of CSAM and CSEA is indicated by its appeals data, as users can submit appeals against content removals if they believe TikTok has made a mistake. Under the 'Safety & Civility - Youth Exploitation & Abuse' policy, only 64,804 of the 1,541,762 total videos removed were successfully appealed and reversed. Under 'Youth Safety & Well-Being - Youth Exploitation & Abuse', only 27,552 of the 1,289,147 total videos removed were successfully appealed and reversed.

4. Key stakeholder engagement:

For detailed information on how TikTok has engaged with external stakeholders to inform its approach to risk identification and risk mitigation, please refer to Annex 4. This Annex provides an in-depth overview of the collaborative efforts and consultations that have shaped TikTok's strategies in these areas.

5. Prioritisation:

TikTok has reported above on its progress in the last year, and below it states key additional actions for the year ahead. TikTok has closely considered its risk environment and the inherent and residual risk discussed above. TikTok plans to devote extra resources to combatting CSAM and CSEA in the year ahead and as a result, CSAM and CSEA remains a Tier 1 risk. As noted in Year 1, TikTok will continue to keep its prioritisation under review.

6. Planned further mitigation effectiveness improvements:

- **Article 35(1)(c), Holistic account content moderation**: TikTok has updated its moderation workflows and policy guidelines to enable holistic account review for moderators, with a focus on **'Fetishising Youth'** policy to detect borderline fetish collector accounts and **'Youth Sexual Solicitation'** policy to detect solicitation signals that cut across features. TikTok is enhancing its detection of CSAM by developing new models to better identify AIGC CSAM, including still images in addition to videos;
- **Article 35(1)(c), Safety strategies targeting suspicious predatory behaviour**: TikTok is developing a policy framework outlining suspicious predatory signals and product safety strategies to mitigate interactions between potential predators and minors, and among potential predator networks;
- **Article 35(1)(f), Increase coverage of Third Party APIs and Keyword Lists**: TikTok already

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014634

ingests keyword and hash lists and uses tools provided by entities including IWF, NCMEC, Google and Microsoft. We plan to integrate these signals into more feature and detection surfaces across the company in 2024;

- **Article 35(1)(c), Improve operational efficiencies to manage increased volume in reporting**: TikTok plans to launch a Child Safety Team operational effectiveness plan to establish a follow-the-sun model for cross regional specialist queue support. 'Follow-the-sun model' refers to a global workflow in which content can be distributed across locations in different time zones during working hours. This will reduce the time taken to review content and ensure high risk cases are prioritised and actioned within an appropriate time frame; and
- **Article 35(1)(e), Adapting advertising systems**: TikTok commits to establishing a robust risk governance process for CSAM in relation to its ads products. TikTok will also continue to monitor the robustness of its ad policies with respect to CSAM to ensure that they cover emerging CSAM risks identified in internal risk assessments, including new forms of online abuse that could fall within the scope of CSAM. Last, TikTok will continue to provide training to relevant teams, helping them to identify violative content under any updated and enhanced ad policies regarding CSAM.

10. ILLEGAL CONTENT: RISKS OF GENDER-BASED VIOLENCE CONTENT

1. Description of the risk:

a. Description of the risk from Year 1

- TikTok understands that neither the DSA nor EU law contains a single or comprehensive definition of gender-based violence. TikTok therefore understands gender-based violence to be the perpetration, support or incitement of 'any type of harm [...] against a person or group of people because of their factual or perceived sex, gender [...] and/or gender identity' ('**GBV**');[18] [19]
- Directive (EU) 2024/1385 on combating violence against women and domestic violence defines 'violence against women' as 'all acts of gender-based violence directed against a woman or a girl because she is a woman or a girl or that affect women or girls disproportionately, that result in or are likely to result in physical, sexual, psychological or economic harm or suffering, including threats of such acts, coercion or arbitrary deprivation of liberty, whether occurring in public or in private life'. This definition is consistent with, yet narrower than, TikTok's interpretation of GBV set out above, as violence against women is only one form of GBV. TikTok has therefore assessed the nature of GBV risks on its Platform using its definition above;
- In the specific context of the Platform, the risk relating to GBV may involve users attempting to share or disseminate content depicting or involving the following types of behaviour on or through video, livestream, comments and in profile information on the Platform:
    - non-consensual sexual acts that are real or fictional, including rape, molestation, and non-consensual touching;

[18] The Council of Europe's definition is based on the Explanatory Report to the Convention on preventing and combating violence against women and domestic violence.

[19] Note that TikTok considers violative content related to hate speech against people due to their sexual orientation in the separate Hate Speech section of this Report.

- promoting violence, exclusion, segregation, discrimination, and other harms on the basis of a protected attribute (i.e. such as gender, gender identity or sex);
- threatening or expressing a desire to cause physical injury to a person or a group (where gender is a relevant factor); and
- degrading someone or expressing disgust on the basis of their personal characteristics or circumstances, such as their physical appearance, intellect, personality traits, and hygiene (where gender is a relevant factor) (together, **'GBV Content'**).

- In addition, TikTok notes that Article 21 of the Charter prohibits discrimination based on sex and sexual orientation, and that Article 23 enshrines the right to equality between men and women. TikTok further notes that GBV, in particular violence against women and domestic violence, violates fundamental rights such as the right to human dignity, the right to life and integrity of the person, the prohibition of inhuman or degrading treatment or punishment, the right to respect for private and family life, personal data protection, and the rights of the child, as enshrined in the Charter.[20]

b. Risk Identification: changes to the risk profile since Year 1

In its Year 1 Risk Assessment, TikTok classified GBV Content as a **Tier 1 priority** in its priority tiering system.

This section summarises additional risks identified in Year 2. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk detection processes, and extensive consultations with external stakeholders.

Summary of risks identified in Year 1

TikTok identified the following inherent risks as part of its Year 1 Risk Assessment:

- **Sharing or dissemination of GBV Content**: Users may attempt to use the Platform to share GBV Content;
- **Content moderation systems:** Given that GBV can be nuanced and highly localised across regional, cultural and linguistic differences in the EU, there is a risk that TikTok's automated detection systems may not keep pace with evolving vocabulary and GBV behaviours, such as coded and localised gendered slurs; and
- **Intentional manipulation of TikTok via impersonation**: Bad actors may use fake accounts to harass and silence victims, particularly in the context of Sexual Exploitation and Abuse. They may impersonate individuals, including victims, to post harmful content and discredit victims by spreading rumours or questioning their experiences. Impersonators may also solicit or share explicit content under the guise of someone the victim trusts, worsening the harm of GBV.

Additional risks identified in Year 2

In Year 2, TikTok identified the following additional risk areas, which it considers to be reflective of the dynamic and evolving nature of GBV Content.

- **AI-generated image based sexual abuse:** Image Based Sexual Abuse (**'IBSA'**) is the creation, manufacture, or distribution of nude, partially nude, or sexually explicit content without the

[20] Proposal for a Directive on combating violence against women and domestic violence, Recital 8.

consent of the person in the content, for the purpose of sexualising their body, or portraying them in a sexual manner. TikTok considers the harm of IBSA to be the same whether it includes real or manipulated media given the psychological impact to the survivor. The risk involves possessing, distributing, or providing instructions on how to create or access AIGC or altered intimate images that were created or distributed without consent;

- **Sextortion**: TikTok identified sextortion as an evolving risk due to NCMEC reports and escalations on the Platform in the sextortion space in the past year, in particular relating to the 'Yahoo Boys' sextortion group;[21] and
- **Sexual harassment of female public figures**: TikTok identified an increased risk of videos and comments which bully or harass female candidates, journalists, and other public figures during elections in response to NGO reports, and external engagements. The risk is particularly relevant in Year 2 as almost half of the world's population is voting in national elections in 2024[22].

c. Inherent Risk in Year 2

For a detailed analysis on how TikTok assessed the baseline severity and probability for GBV in Year 1, please refer to TikTok's Year 1 Systemic Risk Assessment Report.

In Year 2, TikTok assesses the overall severity of GBV to be 'Material'. TikTok assesses the probability of GBV in Year 2 to be 'Possible'. TikTok has therefore assessed the inherent risk (that is, the risk without any mitigations in place) of GBV being shared on the Platform in Year 2 to be 'Moderate'.

2. Mitigation Measures:

Please refer to the Year 1 Risk Assessment Report for detailed information on TikTok's system of mitigations/controls (under DSA Art. 35 (1)(a)-(k)) that applied when its designation as a VLOP came into force. That Year 1 Report identified further mitigation improvements, progress on which (and any other such mitigations) is listed below. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk monitoring and detection processes, and extensive consultations with external stakeholders.

Implementation of Additional Mitigation Effectiveness Measures in Year 2	
MEASURES IN ACCORDANCE WITH DSA ART. 35(1) A-K	**DESCRIPTION**
(b) Adaptation of terms and conditions	In April 2024, TikTok launched an update to the Adult Physical and Sexual Abuse section in its Community Guidelines. The guidelines enhance accessibility and transparency, clarifying further that TikTok does not allow showing, promoting, or engaging in adult sexual or physical abuse

[21] Collection of cybercriminals who are the main drivers of financial sextortion on social media platforms. https://www.wired.com/story/yahoo-boys-scammers-facebook-telegram-tiktok-youtube/ https://longreads.com/2023/07/11/inside-the-world-of-nigerian-yahoo-boys-atavist-excerpt/

[22] The sentence '*half of the world's population voting in national elections in 2024*' has been frequently used by the media to define the magnitude of the 2024 super cycle elections (Source: Reuters, The New Yorker).

	or exploitation. This includes non-consensual sexual acts, IBSA, sextortion, physical abuse, and sexual harassment. The Community Guidelines clearly articulate that the Platform is committed to providing a space that embraces gender equality, supports healthy relationships and respects intimate privacy.
(c) Adaptation of content moderation processes	To continue to mitigate the emerging risk of IBSA, TikTok uses a hash-matching system created in partnership with StopNCII, to proactively identify and remove potential non-consensual intimate imagery content on the Platform. TikTok continues to update keywords to support proactive and automated detection of GBV Content through consultations with external experts and government agencies to ensure keyword lists are up to date. For detailed information on how TikTok has engaged with external stakeholders to inform its approach to identifying and mitigating the risk of Gender Based Violence Content please refer to Annex 4.
(i) Awareness measures	A consultation with Dr. Indira Henard, Executive Director of the DC Rape Crisis Center and Board Member of the National Alliance to End Sexual Violence, resulted in a review of the sexual abuse focused page of the Safety Centre. TikTok recognises that there is language regarding GBV that is not centred on survivors' experiences which can exacerbate trauma and hinder recovery. As a result, TikTok updated its language to be more survivor-centred. For detailed information on how TikTok has engaged with external stakeholders to inform its approach to identifying and mitigating the risk of Gender Based Violence Content please refer to Annex 4.

Case Study: Sextortion

Sextortion risks have evolved in the past year in the way it manifests online. Sextortion is a type of blackmail in which the perpetrator uses threats and intimidation to extort money from their victim. In most cases, the threat is related to the distribution or publication of intimate content of the victim.

In partnership with **Toluna,** TikTok's User Research and Product teams created a survey in November 2023, to get user feedback on their experiences with sextortion across multiple social media platforms. Before this study, TikTok's Risk Analysis team also ran internal checks to analyse the existence of sextortion on TikTok.

This study concluded that TikTok is generally not the Platform of choice for exploitation predators, who are drawn to other social media platforms that allow predators to message users that do not follow them.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014638

Through this partnership with **Toluna,** TikTok gained a more robust understanding of the nature of this risk on its Platform and was able to confirm that existing mitigations for this risk are performing well, including moderation of videos and text to recognise sextortion behaviour, as these tools are based on banks of similar content to capture previously known violations. Of the social media platforms listed in the survey, TikTok has the lowest rate of sextortion.

3. Residual Risk in Year 2:

Following an assessment of the effectiveness, reasonableness and proportionality of TikTok's mitigations relevant to the systemic risk of GBV, TikTok has assessed residual risk to be 'Moderate' in Year 2. TikTok commits to continuing to expand and invest in mitigation measures to identify and mitigate GBV on the Platform.

This assessment was conducted on the basis of qualitative and quantitative insights, including data on TikTok's enforcement of its Community Guidelines in the EU for the period of Q3 2023 to Q2 2024. Under the Community Guidelines policy of 'Sexual Exploitation & Gender-Based Violence', TikTok removed 617,156 total videos in the EU, with 426,161 detected and removed proactively, and 136,813 removed without any views. This high proportion of proactive removal suggests that TikTok's content moderation systems remain effective in detecting and removing GBV Content on the Platform, helping to mitigate the systemic risk.

This assessment of the effectiveness of TikTok's proactive detection systems is further informed by the following data on user reports, which serve as a useful indication of the volume of GBV Content that is not detected and removed by TikTok's proactive systems. User reports accounted for a minority of GBV Content detected and removed by TikTok, at 190,995 of the 617,156 total videos detected and removed under 'Sexual Exploitation & Gender-Based Violence'. TikTok remains committed to continuously working to improve and iterate its proactive detection capabilities, while noting that GBV Content is a rapidly evolving risk area, with nuance across regions and languages.

TikTok's user reporting tools are an important component of its content moderation process, ensuring that GBV Content is detected and removed, including new and evolving forms of GBV. The data demonstrates that when TikTok receives user reports of violative GBV Content, it is actioned efficiently. Under 'Sexual Exploitation & Gender-Based Violence', 166,689 of the 190,995 total videos removed following user reports were removed within two hours of receiving the report. Collectively, TikTok's proactive detection and reporting tools have worked effectively to ensure that 509,501 of the 617,156 total videos removed under 'Sexual Exploitation & Gender-Based Violence' were removed within 24 hours of upload.

The effectiveness, reasonableness, and proportionality of TikTok's moderation of GBV Content is indicated by its appeals data, as users can submit appeals against content removals if they believe TikTok has made a mistake. Under the 'Sexual Exploitation & Gender-Based Violence' policy, only 35,804 of the 617,156 total videos removed were successfully appealed and reversed.

4. Key stakeholder engagement:

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014639

For detailed information on how TikTok has engaged with external stakeholders to inform its approach to risk identification and risk mitigation, please refer to Annex 4 This Annex provides an in-depth overview of the collaborative efforts and consultations that have shaped TikTok's strategies in these areas.

5. Prioritisation:

TikTok has reported above on its progress in the last year, and below it states key additional actions for the year ahead. TikTok has closely considered its risk environment and the inherent and residual risk discussed above. TikTok plans to devote extra resources to combating GBV Content in the year ahead and as a result it remains a Tier 1 priority. As noted in Year 1, TikTok will continue to keep its prioritisation under review.

6. Planned further mitigation effectiveness improvements:

- **Article 35(1)(b), GBV policies:** TikTok plans to launch new policies and enforcement measures targeting GBV Content related to the promotion of exploitation and abuse, such as intimate partner violence. This includes addressing content that normalises unhealthy behaviours in relationships, such as physical and verbal abuse, and coercive control. The policies will also cover fictional representations of exploitation and abuse, as these can contribute to normalising this form of GBV;
- **Article 35(1)(b), AIGC IBSA:** TikTok will update its policy coverage and enforcement of GBV Content to further mitigate the risk of AIGC IBSA that occurs on the Platform or that promotes or directs to off-platform sites. TikTok will also clarify its policy on the harassment of female public figures by creating specific moderator guidelines to help identify sexual harassment, particularly against female politicians; and
- **Article 35(1)(i), User education:** TikTok will provide new messaging safety tips to users to enhance user education in order to mitigate the risk of sextortion. This will include redirecting users to relevant third party support resources.

11. YOUTH SAFETY: RISKS RELATED TO AGE APPROPRIATE CONTENT AND ONLINE ENGAGEMENT

1. Description of the risk:

a. Description of the risk from Year 1

- TikTok is a platform accessible to minors and therefore gives careful and appropriate consideration to the risk encompassed by Article 28(1) DSA, which requires it to put in place appropriate and proportionate measures to ensure a high level of privacy, safety and security for minors on its service. TikTok's minimum age requirement is 13 years old. TikTok uses the term **'Underage Users'** to refer to minors under the age of 13 who may attempt to access the Platform. **'Younger Users'** and **'Youth'** refer to TikTok users with a declared age of between 13 to 17 years

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014640

old inclusive. However, TikTok is not specifically aimed at or predominantly used by minors aged under 18 years old in Europe; and

- TikTok acknowledges that a balance must be found between the design of the measures to address risks to Younger Users and the Fundamental Rights of legitimate users of any age. Those rights include the right to freedom of expression, which includes the right of users to express themselves freely on the Platform, receive and impart information and ideas to aid their individual learning and development and to exercise autonomy. Accordingly, TikTok seeks to avoid unnecessary and disproportionate impacts on such legitimate users in the design of its Platform and the specific Youth Safety measures summarised in this Report;
- TikTok recognises the following Youth Safety risks:
    - **'Content Risk'**, also referred to as risks related to '**Age Appropriate Content**': Younger Users might access or view content on the Platform that is not age-appropriate. Related risks may involve negative effects on physical or mental well-being; and
    - **'Conduct Risk**' and **'Contact Risk'** (collectively, risks related to '**Online Engagement**'): In creating and posting content on the Platform, or by engaging with content posted by others, Younger Users may engage in inappropriate behaviour or potentially encounter inappropriate behaviour from other users, such as inappropriate comments, bullying or behaviour amounting to child sexual exploitation.[23] TikTok notes that this risk only arises in relation to users who are active, rather than passive (i.e., those that only watch content and take no action in relation to it).

b. Risk Identification: changes to the risk profile since Year 1

In its Year 1 Risk Assessment, risks related to Age Appropriate Content and Online Engagement were assessed together with risks related to Age Assurance as 'Online Protection of Minors and Associated Risks'. Based on this assessment, TikTok classified risks related to Youth Safety as a **Tier 1 priority** in its priority tiering system.

Summary of risks identified in Year 1

TikTok identified the following inherent risks as part of its Year 1 Risk Assessment:

- **Creating and publishing content ('Conduct Risk')**: Younger Users may post content without understanding the consequences, including violating Community Guidelines, sharing illegal content, or private information, leading to loss of privacy, embarrassment, or distress;
- **Proactively engaging with content ('Content Risk')**: Younger Users may engage with other creators' content, including commenting or using Stitch or Duet tools, potentially leading to risky or harmful behaviour;
- **Others engaging with Younger Users and their content ('Contact Risk')**: Published content can attract interactions from other users, such as sharing, commenting, or using Stitch or Duet tools, which may result in inappropriate behaviour towards youth, causing embarrassment or distress; and
- **Potential impacts to well-being**: Younger Users may experience negative effects to their well-being, such as spending excessive time on the Platform or feeling pressure to post content and receive likes, impacting their self-esteem and overall well-being.

[23] CSAM and CSEA risks are dealt with in the above section of this Report on risks of child sexual abuse material and child sexual exploitation.

Additional risks identified in Year 2

In Year 2, TikTok identified the following additional risk areas, which it considers to be reflective of the dynamic and evolving nature of f risks related to Age Appropriate Content and Online Engagement.

- **Screen time:** The risk to mental and physical well-being of excessive use of the Platform was highlighted in its Year 1 Risk Assessment. In Year 2, TikTok has identified some specific risk factors that it considers to be relevant:
    - **Online platforms' ease of use**: The ease of using an online platform may make it difficult for Younger Users to intentionally take breaks as they may struggle with self-regulation and impulse control due to their developmental stage; and
    - **Incentives to use an online platform and engage with more content or users**: For Younger Users, who may be focused on peer acceptance and feedback as part of their identity development, the desire for positive social rewards (e.g. likes on a video) may further drive their online engagement and screen time;[24]
- **Risks related to AIGC***:* AIGC is an emerging area with certain specific concerns related to Youth Safety:
    - **Additional content harms**: Younger Users may create and publish AIGC featuring other individuals that is inappropriate or in violation of our Community Guidelines, which could lead to distress, embarrassment, loss of privacy or societal harm if this content is viewed by other users. They may also be misled by AIGC, making Younger Users potentially more susceptible to harm from misinformation, more susceptible to misunderstanding satire or humour, and/or more susceptible to adopting unrealistic standards for personal physical appearance and video quality;
    - **Increased engagement with Younger Users' content**: AIGC may increase engagement with Younger Users' content by making it more appealing or interesting. This increased attention might overwhelm Younger Users, who may not be equipped to handle it safely; and
    - **Well-being impact and social comparison**: Early research indicates there are risks to Younger Users' well-being associated with AIGC, including content created using off-platform AIGC tools, in particular around self-esteem, body image, and mental health. A lack of agreed knowledge and industry best practices to mitigate these effects is a risk factor.
    - For more detailed information on TikTok's approach to mitigating risks associated with AIGC, please refer to *Priority Considerations Across all Modules*.

c. Inherent Risk in Year 2

For a detailed analysis on how TikTok assessed the baseline severity and probability for Online Engagement and Age Appropriate Content in Year 1, please refer to the 'Online Protection of Minors' section in TikTok's Year 1 Systemic Risk Assessment Report.

[24] TikTok relied on expert and academic research to assess this risk, including research produced by the Digital Wellness Lab, American Association of Pediatrics, National Sleep Foundation, Pew Research Center, American Psychological Association, and Common Sense Media.

In Year 2, TikTok assesses the overall severity of risks related to Online Engagement and Age Appropriate Content to be 'Material'. TikTok assesses the probability of risks related to Online Engagement and Age Appropriate Content to be 'Likely' in Year 2. Taking into account the severity and probability in Year 2, TikTok has assessed the inherent risk of risks (that is, the risk without any mitigations in place) related to Online Engagement and Age Appropriate Content in Year 2 to be 'Material'.

2. Mitigation Measures:

For detailed information on TikTok's system of mitigations/controls (under DSA Art. 35 (1)(a)-(k)) that applied when its designation as a VLOP came into force, please refer to the Year 1 Risk Assessment Report. That Year 1 report identified further mitigation improvements, progress on which (and any other such mitigations) is listed below. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk monitoring and detection processes, and extensive consultations with external stakeholders.

Implementation of Additional Mitigation Effectiveness Measures in Year 2	
MEASURES IN ACCORDANCE WITH DSA ART. 35(1) A-I	**DESCRIPTION**
(a) Adaptation of feature or platform design	As mentioned in the Year 1 Report, a key mitigation TikTok implemented was screen time management features - including 60 minute a day default screen time limits for Younger Users and availability of that feature for all users. In the last year, TikTok further integrated and expanded its screen time management features. These features are also now available on the web app version of the Platform (i.e., TikTok.com). TikTok has also implemented delayed screen time interventions in order to increase the interventions' effectiveness. Screen time interventions are often triggered while users are watching a video, which forces the disruptive decision of finishing what they are watching or adhering to their screen time limits. Some users will want to finish watching the video they're currently viewing and will ignore the pop-up. Consequently, TikTok has launched a delay, which prevents the app from showing screen time management interventions until just after a user navigates away from the video, or has watched the video for 60 seconds. If they are not currently watching a video (via Home tab *or* Search), then there's no change to the trigger logic. This is intended to operate so that users are more likely to pay attention to the screen time intervention. TikTok has also launched the STEM Feed, a specialised feed dedicated to showcasing content related to Science, Technology, Engineering, and Math. The STEM Feed is enabled by default for all Younger Users, while others can activate it through content preferences. It's designed as a

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	space where users can discover reputable, educational content outside of the FYF. As mentioned in Year 1 as a key mitigation improvement, TikTok also expanded its Content Classification system (which rates and then displays UGC based on maturity of the user, based on declared age) to include LIVE content. This is used to prevent livestream content with overly mature themes from reaching Younger Users. Since the Year 1 Risk Assessment Report, a new registration pop-up has been introduced for new 16-17 year olds (and a pop-up was presented to existing 16-17 year olds with public accounts) to inform them about the implications of public accounts, with the private account option pre-selected and requiring the 16-17 year old user to make a selection. If no selection is made, the 16-17 year old user's account is defaulted to private.
(b) Adaptation of terms and conditions	In May 2024, TikTok updated its Community Guidelines to address emerging and ongoing risks: • Youth Safety and Well-being: TikTok updated its Youth Safety and Well-being policies around appropriate and inappropriate content, including detailed enforcement outcomes ('not allowed', 'restricted', 'FYF ineligible') and updated key terms related to Youth Safety such as CSAM, grooming, body exposure, and disordered eating; • AIGC: TikTok updated its AIGC guidelines and policies to remove any AI-generated visual or audio representations of younger users if the platform becomes aware of such instances. • Harmful Body Idealisation: In June 2024 TikTok implemented a Harmful Body Idealisation policy which makes ineligible for the FYF content that promotes idealised body aesthetics associated with harmful weight management behaviours, such as disordered eating and anabolic steroid use. TikTok may undertake proactive ad hoc searches for this type of content from time to time to enforce this policy.
(c) Adaptation of content moderation processes	In addition to the Year 1 content moderation mitigations still in place, TikTok has implemented an early detection model. In 2023, TikTok updated the early detection model that it uses to identify and stop potentially dangerous trends and challenges. TikTok also established a 'Known Dangerous Variants' database, which is accessible by its moderators, and which documents frequently (re)surfacing dangerous trends and challenges. Moderators can then refer to this database in their content moderation decisions for guidance.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014644

(d) Adaptation of algorithmic systems	As mentioned in Year 1 as a key mitigation improvement, TikTok has improved its existing Content Classification process. In addition to (a) above, it has also done so by refining its approach to better prioritise accuracy of the automatic labelling of mature content , minimise views of violative content, and remove egregious content quickly. TikTok has developed and integrated a Content Level 1 system for all content channels accessed by logged-out users in the TikTok app. Content Level 1 refers to content that is appropriate for general audiences to view. This aims to ensure that Younger Users using the Platform without logging in are only exposed to Age Appropriate Content. TikTok's Restricted Mode feature, available to all users and also as part of Family Pairing, provides users with the ability to restrict their viewing to Content Level 1 content. TikTok continues to invest in strengthening its content classification system to provide an age-appropriate experience for Younger Users. TikTok has also changed how certain categories of content appear on the FYF by updating its models to prevent the display of concentrated content for topics such as dietary discussions, weight loss, sadness, grief, loneliness, and hopelessness.
(e) Reinforcing risk detection measures	As mentioned in Year 1 as a key mitigation improvement, TikTok continues to monitor existing metrics, model efficacies, and related safety features by sampling videos viewed by teenagers and reviewing them manually to understand if teenagers are viewing content that is appropriate for them to view based on Content Classification Policies. This sampling and reviewing process occurs regularly and is key to detect emerging risks impacting Younger Users.
(i) Awareness measures	TikTok has launched a global Youth Council to reinforce its commitment to enhancing safety measures for teens on the Platform and to hear from teens directly. This initiative responds to recent global research indicating that teens and their parents want more collaborative opportunities with online platforms. Created in partnership with Praesidio Safeguarding, the Youth Council consists of 15 teens from diverse backgrounds and countries, including the United States, United Kingdom, Brazil, Indonesia, Ireland, Kenya, Mexico, and Morocco. The group, which first convened in December 2023 and met again in February 2024 with TikTok Chief Executive Officer Shou Chew, has prioritised teen well-being and inclusion for the year. The Youth Council's insights help TikTok to enhance resources for Younger Users and to refine its risk detection measures and enhance the overall user experience for young people. TikTok is updating its Youth Guide (also known as/now renamed Youth Portal) to include new safety features, based on feedback from the Youth Council on reporting and blocking, and is introducing new GIFs to help Younger Users locate these features. Members of the Youth Council provided input and asked TikTok to share more information about

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014645

	reporting and blocking content and users to better understand what happens after a report is made. For detailed information on how TikTok has engaged with external stakeholders to inform its approach to identifying and mitigating risks related to Age Appropriate Content and Online Engagement please refer to Annex 4.

3. Residual Risk in Year 2:

Following an assessment of the effectiveness, reasonableness and proportionality of TikTok's mitigations relevant to the systemic risk related to Age Appropriate Content and Online Engagement, TikTok has assessed residual risk to be 'Moderate' in Year 2. TikTok commits to continuing to expand and invest in mitigation measures to identify and mitigate risks related to Online Engagement and Age Appropriate Content on the Platform.

This assessment was conducted on the basis of qualitative and quantitative insights, including data on TikTok's enforcement of its Community Guidelines in the EU for the period of Q3 2023 to Q2 2024. Under the Community Guidelines policy of 'Youth Safety and Well-Being', TikTok restricted 8,155,040 total videos in the EU, with 7,967,430 detected and restricted proactively, and 5,871,273 restricted without any views. This high proportion of proactive removal suggests that TikTok's content moderation systems remain effective in detecting and restricting content endangering Younger Users on the Platform, helping to mitigate the systemic risk.

This assessment of the effectiveness of TikTok's proactive detection systems is further informed by the following data on user reports, which serve as a useful indication of the volume of content contributing to this systemic risk that is not proactively detected by TikTok's proactive systems. User reports accounted for a minority of content detected and restricted due to systemic risks related to Age Appropriate Content and Online Engagement, at 187,610 of the 8,155,040 total videos detected and restricted under 'Youth Safety and Well-Being'. TikTok remains committed to continuously working to improve and iterate its proactive detection capabilities, while noting that risks related to Age Appropriate Content and Online Engagement are rapidly evolving, with nuance across regions and languages.

TikTok's user reporting tools are an important component of its content moderation process, ensuring that content endangering Younger Users is detected and restricted, including new and evolving risks related to Age Appropriate Content and Online Engagement. The data demonstrates that when TikTok receives user reports of violative content under the 'Youth Safety and Well-Being' policy, they are actioned efficiently. Under 'Youth Safety and Well-Being', 158,067 of the 187,610 total videos restricted following user reports were removed within two hours of receiving the report. Collectively, TikTok's proactive detection and reporting tools have worked effectively to ensure that 7,132,887 of the 8,155,040 total videos restricted under this policy were restricted within 24 hours of upload.

The effectiveness of TikTok's moderation of risks related to Age Appropriate Content and Online Engagement is also indicated by its appeals data, as users can submit appeals against content restrictions if they believe TikTok has made a mistake. Under the 'Youth Safety and Well-Being' policy, only 335,486 of the 8,155,040 total videos restricted were successfully appealed and reversed.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014646

4. Key stakeholder engagement:

For detailed information on how TikTok has engaged with external stakeholders to inform its approach to risk identification and risk mitigation, please refer to Annex 4. This Annex provides an in-depth overview of the collaborative efforts and consultations that have shaped TikTok's strategies in these areas.

5. Prioritisation:

TikTok has reported above on its progress in the last year, and below it states key additional actions for the year ahead. TikTok has closely considered its risk environment and the inherent and residual risk discussed above. TikTok plans to devote extra resources to mitigating risks related to Age Appropriate Content and Online Engagement in the year ahead and as a result it is a Tier 1 priority. As noted in Year 1, TikTok will continue to keep its prioritisation under review.

6. Planned further mitigation effectiveness improvements:

- **Article 35(1)(j), Screen management:** TikTok plans to actively explore and roll out new product features that will provide additional functionality to help teens better manage screen time while maintaining a high-quality user experience.
- **Article 35(1)(j), Family pairing**: TikTok plans to enhance the Family Pairing feature by providing parents with greater visibility over their teens' activities and access on TikTok, while respecting teens' rights;
- **Article 35(1)(j), Controls and transparency**: TikTok will continue enhancing the non-personalised experience, keyword filtering, and researcher tools, developing a robust product roadmap for user-facing algorithm control features; and
- **Article 35(1)(c), Content Levels:** TikTok also plans to enhance the content levels across the Platform for its Younger Users by continuing to scale classification from what was originally a manual labelling approach to a model-driven automatic labelling approach, improving the detection and filtering of age inappropriate content. TikTok plans to pilot this expansion outside the EU and then to later expand it to users in the EU;

12. YOUTH SAFETY: RISKS RELATED TO AGE ASSURANCE

1. Description of the risk:

a. **Description of the risk in Year 1**

TikTok understands 'Age Assurance' in the context of Article 34(1) DSA and the previously addressed risks to Younger Users related to Age Appropriate Content and Online Engagement. While Younger Users are entitled to freedom of expression, access to information and other fundamental rights associated with use of the Platform, TikTok considers it reasonable and proportionate to prohibit access for Underage Users and permit Platform access for Younger Users while restricting their access to

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014647

specific features of Platform experiences. 'Age Assurance' refers to the tools and technologies employed by TikTok to prohibit or restrict Youth access to the Platform, and thereby, help protect Underage Users from the risks associated with Online Engagement and Age Appropriate content. TikTok's approach is consistent with relevant international legal frameworks relating to the protection of child safety (e.g. UN Convention on Rights of the Child), which recognise the unique vulnerabilities of children and teenagers.

This module addresses the systemic risk that (1) Underage Users may mis-state their age at registration and that they may gain access to the Platform as a result; and (2) that Younger Users may mis-state their age at registration and not receive an age appropriate experience on the Platform, together the **Mis-Stated Age Risk**. Additionally, this risk includes the possibility of inaccurate age determinations at other stages of the Age Assurance process, such as in underage moderation and appeals.

b. Risk Identification: changes to the risk profile since Year 1

In its Year 1 Risk Assessment, risks related to Age Assurance were assessed together with risks related to Age Appropriate Content and Online Engagement as 'Online Protection of Minors and associated risks'. Based on this assessment, TikTok classified risks related to Youth Safety as a **Tier 1 priority** in its priority tiering system.

This section summarises additional risks identified in Year 2. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk detection processes, and extensive consultations with external stakeholders.

Summary of risks identified in Year 1

TikTok identified the following inherent risks as part of its Year 1 Risk Assessment:

- **Mis-stated age**: Underage Users may attempt to gain access to the Platform, and Younger Users may attempt to avoid an age-appropriate experience by mis-stating their age at the point of registration;
- **Parents or guardians misunderstanding Age Assurance**: Parents or guardians may not appreciate or understand the minimum age requirement, and age-appropriate restrictions, and support or help their child to bypass age assurance measures (e.g. through appeals); and
- **Unreasonable and disproportionate impacts on legitimate users**: Age Assurance tools and processes may not adequately and proportionately protect other fundamental rights, such as adult users' rights to freedom of expression (e.g. erroneous removal of adult users as suspected underage accounts).

Additional risks identified in Year 2

In Year 2, TikTok identified the following additional risk areas, which it considers to be reflective of the dynamic and evolving nature of risks related to Age Assurance.

- **Mis-Stated Age Risk in moderation processes**: When TikTok detects potential mis-stated age accounts, human moderators are tasked with identifying the user's age, however individual cases can occasionally be challenging for moderators to discern. There is therefore a risk of harm to the rights, particularly freedom of expression, of account owners if their age is incorrectly

assessed as being underage, and their account is removed from (or restricted on) the Platform (though users have the right to appeal); and

- **Use of AIGC to circumvent the age appeals process**: Users can appeal TikTok's decision to remove them from the Platform on the basis that it has concluded that they are an Underage User. This process involves using a choice of methods to provide proof of age. There has not been any reported instance of AIGC being used to intentionally create fake appeal data in order for a user to appear older than they are. However, human moderators have reported difficulty in identifying age signals when particular AI features are used in content reviewed by moderators. AIGC could undermine the effectiveness of enforcement against Mis-Stated Age accounts. Without effective enforcement, Underage Users may gain access to the Platform, and Younger Users may avoid age restrictions.

c. Inherent Risk in Year 2

For a detailed analysis of how TikTok assessed the baseline severity and probability for risks related to Age Assurance in Year 1, please refer to TikTok's Year 1 Systemic Risk Assessment Report.

In Year 2, TikTok assesses the overall severity of risks related to Age Assurance to be 'Material'. TikTok assesses the probability of risks related to Age Assurance in Year 2 to be 'Likely'. Taking into account the changes in severity and probability in Year 2, TikTok has assessed the inherent risk (that is, the risk without any mitigations in place) related to Age Assurance in Year 2 to be 'Material'.

2. Mitigation Measures:

For detailed information on TikTok's system of mitigations/controls (under DSA Art. 35 (1)(a)-(k)) that applied when its designation as a VLOP came into force, please refer to the Year 1 Risk Assessment Report. That Year 1 Report identified further mitigation improvements, progress on which (and any other such mitigations) is listed below. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk monitoring and detection processes, and extensive consultations with external stakeholders.

Implementation of Additional Mitigation Effectiveness Measures in Year 2	
MEASURES IN ACCORDANCE WITH DSA ART. 35(1) A-I	**DESCRIPTION**
(a) Adaptation of feature or platform design	TikTok implemented additional steps to protect its age appeals process against the use of AIGC by Underage or Younger Users to misrepresent their age. If users want to confirm their age through the 'Selfie with ID' option, they must upload a live photo of themselves holding their ID directly from the camera function. This helps to mitigate the risk that users upload manipulated or AIGC images from their photo album. Similarly, if Younger Users want to confirm their age through TikTok's 'photo with a parent/guardian or trusted adult' option, the various requirements would be difficult for AIGC to replicate (i.e. selfie with trusted adult holding paper detailing proof of age, date of birth, and unique 6-digit TikTok code, and the photo must also be taken from the

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014649

	camera function). TikTok's 'credit card authorisation' appeal option for users aged over 18 would be close to impossible for AIGC to replicate. To complement these three existing age appeals options, TikTok implemented an age estimation technology for users aged 18 and over, as mentioned in the Year 1 Report. This new appeal option has been developed by Yoti, a specialised partner that was selected based on its superior technological capabilities, regulatory approval and industry support. Users can provide a selfie taken through a secure image capture function that is sent to Yoti, which uses its technology to conduct an age estimation before returning the result to TikTok and deleting the selfie from their systems. The secure image capture function will prevent users from uploading AIGC or manipulated images that may interfere with age estimation by misrepresenting their age. This tool expands appeal options to users who do not want, or are unable to, use a credit card or ID for age confirmation.
(c) Adaptation of content moderation processes	As mentioned in the Year 1 Report as a planned mitigation, in January 2024, TikTok updated its 'Suspected Underage User' moderation guidance to reflect two distinct policies; 'Suspected U13 user' and 'U13 Age admission'. This change allows moderators to identify visual-based signals and admission-based age (i.e., the user admitting they are under 13) signals separately. Content tagged with one of these two new policy titles will be reviewed by human moderators, with the content removed and account banned if the account is found to be that of an Underage User. Prior to this change, only one policy was applied (i.e. 'Suspected Underage User'), which did not differentiate between visual and admission-based signals. Admission-based signals tend to support a more straightforward assessment of age, and so, providing separate policy titles will allow reviewers to consider the varying accuracy of respective age signals in deciding on enforcement. This policy adjustment supports more accurate enforcement against underage accounts. Additionally, using separate internal tags for 'Suspected U13 Age' and 'U13 Age Admission' means that TikTok can monitor and improve U13 detection in a more granular way. In advance of bifurcating its 'Suspected Underage User' policy title, TikTok launched new moderation guidance that provided additional guidance for moderators to use in identifying visual- and admission-based age signals. This includes guidance on the type of physical features, behaviours and environments that are typical of Underage Asers (i.e. visual-based), and other signals such as users using euphemisms to communicate their age (i.e. admission-based). While these factors have long featured in TikTok's underage moderation, the new guidance reflects advances in TikTok's understanding of the risk, and incorporates trends and tactics observed in Underage Users. This

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014650

	ultimately helps moderators to attach the appropriate policy title under TikTok's new bifurcated policies, supporting appropriate action on underage accounts. In January 2024, TikTok also provided updated and revised guidance to address underage moderation challenges associated with AIGC. Under this new guidance, if a user's face or other attributes are obscured and there is reason to believe that the user is underage, moderators are advised to use extra caution and tag the video as suspected underage. These accounts are subsequently reviewed by account-level moderators who have access to more information (e.g. profile bio, profile photo, more content posted by the account).

3. Residual Risk in Year 2:

Following an assessment of the effectiveness, reasonableness and proportionality of TikTok's mitigations relevant to the systemic risk related to Age Assurance, TikTok has assessed residual risk to be 'Moderate' in Year 2. TikTok commits to continuing to expand and invest in mitigation measures to identify and mitigate risks to Age Assurance on the Platform.

This assessment was conducted on the basis of qualitative and quantitative insights, including data on TikTok's enforcement of its Community Guidelines in the EU for the period of Q3 2023 to Q2 2024.[25] TikTok's mitigations aim to detect and remove Underage Users, and detect and restrict Younger Users. TikTok detected and removed 3,232,672 Underage Users from the Platform in the EU from Q3 2023 to Q4 2023 .

4. Key stakeholder engagement:

For detailed information on how TikTok has engaged with external stakeholders to inform its approach to risk identification and risk mitigation, please refer to Annex 4. This Annex provides an in-depth overview of the collaborative efforts and consultations that have shaped TikTok's strategies in these areas.

5. Prioritisation:

TikTok has reported above on its progress in the last year, and below it states key additional actions for the year ahead. TikTok has closely considered its risk environment and the inherent and residual risk discussed above. TikTok plans to devote extra resources to mitigating risks related to Age Assurance in the year ahead and as a result it remains a Tier 1 risk. As noted in Year 1, TikTok will continue to keep its prioritisation under review.

6. Planned further mitigation effectiveness improvements:

[25] TikTok can provide data on U13 accounts detected and removed for Q3 and Q4 2023, but not the remainder of the Year 2 reporting period. Going forward, TikTok intends to ensure that it can provide this data for the full reporting period, as per the CGER data provided in other risk sections of this report.

- **Article 35(1)(c), Continually improve and monitor quality of moderation of suspected Underage Users and Younger Users:** TikTok will improve visual signals-based guidelines and other instructions provided to moderators when reviewing suspected Underage User and Younger User accounts that reflect Trust & Safety's assessment of risks and recent trends seen in enforcement;
- **Article 35(1)(c), Bias mitigation:** TikTok will strengthen mechanisms to better identify, and reduce, bias in age assurance processes;
- **Article 35(1)(c), Monitoring of age detection models:** TikTok will continue to monitor and assess its age detection models for Underage Users and Younger Users in order to further enhance age assurance processes;
- **Article 35(1)(f), Age-based product mitigations:** TikTok will provide advice to product teams on when to apply age-based mitigations in product development to ensure consistent application, based on Trust & Safety's assessment of age-related risks; and
- **Article 35(1)(j), Expand age assurance vendor and appeal types:** TikTok will evaluate increasing the use of third-party age assurance providers, including in the age appeals process, to improve the user experience while ensuring that user privacy and security is not affected.

13. CIVIC INTEGRITY: RISKS TO ELECTIONS AND CIVIC INTEGRITY

1. Description of the risk:

a. **Description of the risk in Year 1**

b. TikTok understands the risk to be the actual and foreseeable negative effects on election processes and on Civic Integrity arising from the dissemination of verifiably false or misleading content related to an election or other civic events in the EU (such as referenda or a census) (together, '**Election Misinformation**');

c. Election Misinformation risk may arise from attempts to share or disseminate the following content on or through the Platform, whether as short video, comment, livestream or within their profile information:

d. Misleading information about how to vote or register to vote or the qualifications to vote or run in an election;

e. Misleading information about the date of an election or other civic process (e.g. stating that an election is on a later date than it is scheduled for);

f. Misleading information about how to participate in a census or eligibility requirements for participating in a census;

g. Content that advances false claims related to the technical eligibility requirements for current political candidates and sitting elected government officials to serve in office;

h. False claims of election fraud (such as voting machines being tampered with to favour a candidate or political party);

i. Content that falsely claims that an election has been or will be rigged, so the results cannot be trusted;

j. Misinformation or/and conspiracy theories about candidates, candidate impersonation and related issues that may impact civic integrity; or

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014652

k. Synthetic and manipulated media (e.g. modified using model technology) featuring public figures (such as a government official, politician, business leader, or celebrity) may impact Civic Integrity if mis-used for political endorsements or other purposes (TikTok refers to this as '**Synthetic and Manipulated Media**').
l. TikTok also acknowledges that Articles 39 and 40 of the Charter enshrine the rights for every EU citizen to vote and to stand as a candidate at elections to the European Parliament and at municipal elections, respectively, and that such Fundamental Rights may be undermined by Election Misinformation.

Summary of risks identified in Year 1

TikTok identified the following risks as part of its Year 1 Risk Assessment:

a. **Election misinformation:** The risk of attempts to share or disseminate election misinformation that could adversely influence an election, undermine trust in democratic processes, and/or cast doubt on the legitimacy of election outcomes. The specific types of election misinformation identified by TikTok in Year 1 were:
b. Misinformation about how to vote;
c. Misinformation about qualifications to run in an election;
d. Misinformation about the date of an election or other civic processes;
e. False claims about technical eligibility requirements for current political candidates and sitting elected officials;
f. Misleading claims of election fraud; and/or
g. Misinformation and/or conspiracy theories about candidates or candidate impersonation.
h. **Synthetic and manipulated media:** The risks of mis-use of synthetic and manipulated media to depict public figures in misleading ways, such as for political endorsements;
i. **Fake engagement:** The risk of covert influence operations, fake engagement and spam activity that attempts to disrupt election processes, amplify certain viewpoints in elections or sew social disharmony;
j. **Ads:** The risk that ads contain election misinformation.

Additional risks identified in Year 2

In Year 2, TikTok identified the following additional risk areas, which it considers to be reflective of the dynamic and evolving nature of risks to Elections and Civic Integrity.

a. **Geopolitical factors**: The emergence and continuation of two prominent wars (i.e. Ukraine-Russia and Israel-Hamas) and the coincidence of a high number of global elections in 2024 has increased the complexity of identifying and moderating Election Misinformation. These events have necessarily led to a higher volume of political content on the Platform, more complex and nuanced political opinions being shared and new opportunities for bad actors to promote polarising narratives. There is a risk of policy gaps if trends and evolving patterns of discourse outpace policy updates. Without a significant reinforcement of internal resources and processes, added geopolitical complexity creates further risks that Election Misinformation is shared and amplified on the Platform;
b. **Additional potentially harmful types of Election Misinformation**: In the context of heightened geopolitical complexity, TikTok has identified additional types of Election Misinformation that could cause harm on the Platform. This includes misleading information about (or the outright

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014653

denial of) legitimate election outcomes, false claims misrepresenting authoritative sources of civic information (e.g. the text of a parliamentary bill), unverified claims that may harm users' trust in election/civic processes, procedures or outcomes or that of election workers, and manipulated media that falsely appears to come from reputable news organisations. The potential negative impacts on election and civic processes remain the same: adversely influencing public debate, discouraging participation, and damaging trust in democratic processes. The risks were identified through a combination of expert consultations, media monitoring, internal expertise, platform preparations for the 2024 elections, and engagement with relevant communities and events;

c. **Commercial disinformation:** TikTok has also identified commercial disinformation as a new risk in Year 2. Commercial disinformation refers to influence operations with a financial intent or motivation. This includes disinformation-as-a-service sold by third-party vendors, and misleading marketing campaigns presented as UGC. Commercial disinformation can pollute the information environment and reduce user trust in the Platform. This risk was identified through TikTok's collaboration with third party vendors to monitor the emerging capabilities of threat actors, alongside industry reports and other publicly available investigative reporting;

d. **Risks related to AIGC:** The risk of Election Misinformation is further aggravated by emerging and evolving (mis)uses of AIGC. Risks include users sharing synthetic and manipulated media that falsely depicts a public figure in a situation that did not occur, impersonating reputable news sources, and creating content (e.g. AIGC video) that claims to provide 'evidence' that substantiates Harmful Misinformation narratives. Additionally, AIGC potentially makes it easier for bad actors (e.g. coordinated influence operations) to increase the quality and volume of dis- and misinformation production. Internal escalations, consultations with external experts, and TikTok's preparations in advance of the 2024 election cycle, among other detection mechanisms, have also identified the potential risk that AIGC technologies are exploited by covert influence operatives in order to automate malign influence operations. The risks were identified through a combination of expert consultations, media monitoring, internal expertise, platform preparations for the 2024 elections, and engagement with relevant communities and events;

e. **Electoral misconduct:** TikTok has identified electoral misconduct as a new Election Misinformation risk. This risk is that users may share content that promotes or provides instruction on illegal participation and electoral interference, including voter intimidation, and calls for disrupting legitimate election outcomes through violence. Such misconduct can lead to serious harm, including off-platform violence and the subversion of democratic processes. This risk was identified through internal escalations, preparations for the 2024 election cycle, and consultations with experts and civil society groups;

f. **Risks from evolving state-affiliated media activities:** TikTok has identified risks associated with the evolution of state-affiliated media strategies used to interfere with the information environment. This includes efforts to bypass and/or evade TikTok's system for labelling 'State-Affiliated Media' accounts by deploying localised franchises and rebranded entities, using influencers to amplify state-affiliated narratives, and/or using individual media employees to propagate controlled messaging;

g. **Hacked materials:** Hack-and-leak operations may negatively impact or undermine democratic elections by disturbing civic processes, escalating violence, further polarising society, and/or reducing trust in the information ecosystem. In the context of high-priority elections world-wide, TikTok has identified an increased risk of the Platform being used to disseminate hacked materials, with an adverse impact on civic integrity;

h. **Political ads:** TikTok does not allow political advertising in Europe on the Platform. With almost half of the world's population voting in national elections in 2024, bringing a higher volume of potential political content to the Platform, there is a heightened risk that TikTok encounters difficulties in detecting and enforcing against political content in ads; and

i. **Advertising:** TikTok offers 'Promote', a product feature that allows users to purchase advertising inventory for their UGC, ensuring a certain level of traffic for their content. As with other types of advertisements, 'Promote' is subject to TikTok's updated advertising policies, meaning that TikTok's prohibition on political advertising applies to the 'Promote' feature. Although users are informed that using Promote will mean the content must comply with our Community Guidelines and advertising policies, some users still might not consider it advertising when promoting their own UGC.

1. Inherent Risk in Year 2

For a detailed analysis of how TikTok assessed the baseline severity and probability for risks related to Age Assurance in Year 1, please refer to TikTok's Year 1 Systemic Risk Assessment Report.

In Year 2, TikTok assesses the overall severity of risks related to Age Assurance to be 'Material'. TikTok assesses the probability of risks related to Age Assurance in Year 2 to be 'Likely'. Taking into account the changes in severity and probability in Year 2, TikTok has assessed the inherent risk (that is, the risk without any mitigations in place) related to Age Assurance in Year 2 to be 'Material'.

2. Mitigation Measures:

For detailed information on TikTok's system of mitigations/controls (under DSA Art. 35 (1)(a)-(k)) that applied when its designation as a VLOP came into force, please refer to the Year 1 Risk Assessment Report. That Year 1 Report identified further mitigation improvements, progress on which (and any other such mitigations) is listed below. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk monitoring and detection processes, and extensive consultations with external stakeholders.

Implementation of Additional Mitigation Effectiveness Measures in Year 2	
MEASURES IN ACCORDANCE WITH DSA ART. 35(1) A-I	**DESCRIPTION**
(b) Adaptation of terms and conditions	In Year 2, TikTok expanded policy coverage and updated its Community Guidelines in a number of areas, including: • Launch of the 'Unverified Election Claims' Community Guidelines policy which prevents unverified claims about an election from appearing in the FYF (e.g. a premature claim that all ballots have been counted or tallied). This may occur when a new claim has been identified and sent to be fact checked- the label preventing the content from appearing in the FYF is applied until the result of the fact check, when either the label or content is removed; • Launch of the 'Election Misconduct' policy, which prohibits content that encourages or instructs on illegal actions related

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014655

	to elections, including interfering with the electoral process or advocating for the disruption of a legitimate election outcome through unlawful means, such as a coup; • Launch of the 'Distribution of Hacked Materials' policy, which prohibits the sharing of hacked materials that may lead to disturbances in civic processes, escalation of violence, polarisation, or reduced trust in the information ecosystem; • Launch of the 'Misrepresented Civic Sources' policy, prohibits content that severely misleads users about important civic information, which could harm the information ecosystem, undermine public participation in civic processes, and reduce awareness of political issues; and • Updated TikTok's Community Guidelines and advertising policies regarding misleading AIGC and synthetic media to more specifically identify and prohibit specific types and uses of AIGC in UGC and ads (including AIGC that could reduce trust in civic processes and institutions or inflame tensions during crises), ensuring robust policy coverage. These specific Community Guidelines changes are detailed in the cross-risk mitigations outlined for AIGC in the 'Priority Considerations' section.
(c) Adaptation of content moderation processes	TikTok has built a new task prioritisation feature for its misinformation moderation processes. This change means TikTok has begun to move from a 'first-in-first-out' task moderation system to a system that prioritises matters for moderator review based on a machine learning model risk score. The new system will allow TikTok's teams to quickly action the highest risk and most likely violative misinformation first, as opposed to the order in which first appeared in the process to be actioned. Misinformation that falls under TikTok's database of previously fact-checked claims will be immediately actioned, while new instances of potential new Election Misinformation will be sent to a relevant fact-checker partner for review. As mentioned in the Year 1 Report, TikTok has expanded its Fact-Checking Programme to ten partners. These organisations provide fact-checking coverage in 22 official EU languages: Agence France-Presse (AFP), dpa Deutsche Presse-Agentur, Demagog, Facta, Faktograf, Lead Stories, Logically Facts, Newtral, Poligrafo and Science Feedback. For detailed information on TikTok's Fact-Checking Programme, see Annex 5. For detailed information on how TikTok has engaged with external stakeholders to inform its approach to identifying and mitigating risks to Elections and Civic Integrity, please refer to Annex 4. TikTok has reviewed and updated its fact checker ratings. This update includes providing more veracity rating options to fact checkers when they are assessing content, including new and updated rating

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	categories of 'false', 'misleading', 'unsubstantiated', 'inconclusive', 'true' and 'out of scope'. Having a better spectrum of ratings is more reflective of the range of content escalated, and provides options to tag grey-area content. Previously, fact checkers didn't have specific categories to rate content, particularly around evolving events and crises where things may have been rated 'unconfirmed' or 'disputed'. (e.g. 'inconclusive') which resulted in less nuanced assessments and moderation actions. This update has been accompanied by improved definitional language for fact-checkers regarding the scope of each rating and the topical sub-classification options. Given added Election Misinformation risks and complexity associated with the Israel-Hamas and Russia-Ukraine wars, TikTok has launched new guidance to help moderators apply TikTok's policies to more complex and evolving types of violative misinformation content. Previously, moderators did not have sufficient guidance around how to interpret fact checker ratings. TikTok has also continued to routinely update moderator guidelines to account for local languages and political nuances in anticipation of upcoming elections.
(e) Adaptation of advertising systems	TikTok has launched four new, more granular misinformation policies for TikTok's advertising products in Europe on medical misinformation, dangerous misinformation, fake news, and dangerous conspiracy theories. These policies are enforced through proactive keyword-based sweeps of ads uploaded to the Platform. In May 2024, TikTok also updated its policies regarding state-affiliated media accounts to limit their reach on global events and affairs. Now, when TikTok identifies such accounts trying to reach audiences outside their home country, their content will be ineligible for recommendation, meaning the content will not appear in the FYF. Additionally, if these accounts advertise, they will only be allowed to do so within their own country. These changes will be applied to all markets where TikTok's state-controlled media labels are available. In Year 2, TikTok has launched dedicated political advertising content moderation processes in all major monetised EU election markets. Moderators receive market-specific election training, allowing for more nuanced and enhanced enforcement of TikTok's prohibition against political advertising in EU member states. Alongside specialised moderation queues in European markets, TikTok has launched a new enforcement strategy for critical violations of its political advertising policy. This entails a new 'Political Content - Critical' label that suspends advertisers who promote or oppose political causes or outcomes explicitly in ads. The 'Critical' label is distinct from the 'Political Content - General' label, which controls ads that reference political content (e.g. featuring images of candidates). The practical difference is in the punishment applied which is applying the 'General' label to ad content only results in the rejection of the ad,

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014657

	whereas applying the 'Critical' label can also result in an account suspension. Additionally in Year 2, TikTok has developed new automated models that can identify if ads contain any flags, symbols, and/or party logos in order to enhance proactive moderation by the Platform's advertising moderation system.
(f) Reinforcing risk detection measures	In Year 2, TikTok has deployed improved machine learning models to facilitate the effective detection of potential Election Misinformation. The primary detection model now takes into account relevant sub-policies for more accurate identification of violative content (e.g. Election Misinformation, and Medical Misinformation). This means that the model now takes account of different possible types of misinformation and their respective varying characteristics, rather than purely basing its assessment on a binary outcome.
(g) Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21	TikTok remains ready to engage with EU 'trusted flaggers' as they continue to be designated (read section 5) and at the same time, TikTok also continues to receive reports from Community Partners and subject matter experts.
(h) Initiating or adjusting cooperation with other providers of online platforms or of online search engines through the codes of conduct	TikTok has created a rapid response system ('**RRS**') to streamline the exchange of information among civic society organisations, fact-checkers, and online platforms as part of TikTok's commitments as signatories of the Code of Practice on Disinformation ('**CoPD**'). The rapid response system was a time-bound dedicated framework for cooperation among signatories during the 2024 European Parliament elections, allowing non-platform signatories to flag time sensitive content, accounts or trends that may present threats to the integrity of the electoral process. There have been 5 RRS reports received between the beginning of May and the EU election week in early June.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014658

(i) Awareness-raising measures	As part of its ongoing Election Integrity Programme described in Year 1, starting in 2023 and throughout 2024, TikTok implemented in-app Election Centres for each EU member state providing access to authoritative information, in the lead-up to, and during, the 2024 EU Parliamentary Elections. These Centres were developed in collaboration with external stakeholders in a bespoke way for each member state and in the local language. One part of this included collaborating with fact checkers to run media literacy campaigns in EU member states. These campaigns have included embedding a media literacy video focused on critical thinking skills in the EU Election Centre page of each member state. These videos are unique for each member state, with localised detail and context and support users' ability to identify misinformation and manipulated media. For detailed information on how TikTok has engaged with external stakeholders to inform its approach to identifying and mitigating risks to Elections and Civic Integrity, please refer to Annex 4.

3. Residual Risk in Year 2:

Following an assessment of the effectiveness, reasonableness, and proportionality of TikTok's mitigations relevant to the systemic risk of Election Misinformation, TikTok has assessed residual risk to be 'Moderate' in Year 2. TikTok commits to continuing to expand and invest in mitigation measures to identify and mitigate Election Misinformation on the Platform.

TikTok has considered relevant data including data on its enforcement of its Community Guidelines in the EU for the period of Q3 2023 to Q2 2024, data under its Covert Influence Operations reporting, and data collected in relation to the 2024 EU Parliamentary election to support its assessment of the effectiveness of its mitigations against the systemic risk of Election Misinformation.

Under the Community Guidelines policy of 'Civic & Election Integrity', TikTok removed 75,335 total videos in the EU, with 73,930 detected and removed proactively, and 71,841 removed without any views. This high proportion of proactive removal suggests that TikTok's content moderation systems

remain effective in detecting and removing Election Misinformation on the Platform, helping to mitigate the systemic risk.

This assessment of the effectiveness of TikTok's proactive detection systems is further informed by the following data on user reports, which serve as a useful indication of the volume of Election Misinformation that is not proactively detected and removed by TikTok's systems. User reports accounted for a minority of the Election Misinformation detected and removed by TikTok, at 1405 of the 75,335 total videos detected and removed under 'Civic & Elections Integrity'. TikTok remains committed to continuously working to improve and iterate its proactive detection capabilities, while noting that Election Misinformation is a rapidly evolving risk area, with nuance across regions and languages.

TikTok's user reporting tools are an important component of its content moderation process, ensuring that Election Misinformation is detected and removed. The data demonstrates that when TikTok

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014659

receives user reports of violative Election Misinformation content, it is actioned efficiently. Under 'Civic & Election Integrity', 1,383 of the 1405 total videos removed following user reports were removed within

two hours of receiving the report. Collectively, TikTok's proactive detection and reporting tools have worked effectively to ensure that 74,268 of the 75,335 total videos removed under this policy were removed within 24 hours of upload.

The effectiveness of TikTok's moderation of Election Misinformation is indicated by its appeals data, as users can submit appeals against content removals if they believe TikTok has made a mistake. Under the 'Civic & Election Integrity' policy, only 1,885 of the 75,335 total videos removed were successfully appealed and reversed.

TikTok also reports on its efforts to detect and remove covert interference operations ('**CIO**'). TikTok's global data shows that the Platform detected and removed 51 separate new CIO networks from 1 July 2023 to 30 June 2024. These efforts were reinforced by the removal of 23,923 accounts from 1 January to 30 June 2024 associated with previously disrupted networks attempting to re-establish their presence.

TikTok has also collected data regarding user engagement with the Election Centre pages set-up in the context of the EU Parliamentary Elections of June 2024. The Election Centres were linked to users through labels on relevant videos (e.g. videos that engaged with election topics), LIVE streams, and search, with the pages containing authoritative information tailored to each EU member state. From 6 May to 9 June 2024, Election Centre labels received over 4.5 billion views (4,753,349,948) on video and LIVE, with 8,808,681 clicks through to see more information in the Election Centre. Similarly, Elections Centre labels on relevant searches, received over 63 million views (63,613,989) and 46,261 clicks.

4. Key stakeholder engagement:

For detailed information on how TikTok has engaged with external stakeholders to inform its approach to risk identification and risk mitigation, please refer to Annex 4. This Annex provides an in-depth overview of the collaborative efforts and consultations that have shaped TikTok's strategies in these areas.

5. Prioritisation:

TikTok has reported above on its progress in the last year, and below it states key additional actions for the year ahead. TikTok has closely considered its risk environment and the inherent and residual risk discussed above. TikTok plans to devote extra resources to combatting risks of Election Misinformation in the year ahead and as a result it remains a Tier 1 risk. As noted in Year 1, TikTok will continue to keep its prioritisation under review.

6. Planned further mitigation effectiveness improvements:

- **Article 35(1)(c)(i)(f), Elections Integrity Programme**: TikTok will continue to deliver its Elections Integrity Programme for forthcoming elections in EU member states, elevating authoritative sources, civic information and media literacy through in-app product features (e.g. the Election Centre). TikTok will further explore and test ways to bring context to users through contextual labels, especially in relation to unverified election claims. TikTok also plans to update

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014660

its internal risk taxonomy and align its metrics by risk area, enabling more holistic risk detection and mitigation;

- **Article 35(1)(c), Synthetic and manipulated media**: TikTok plans to take steps to strengthen enforcement in relation to Synthetic and Manipulated Media, including around audio and video detection and removal or labelling of manipulated content (including deep fakes);
- **Article 35(1)(e), Adaptation of advertising systems**: TikTok plans to implement a new policy to restrict suspected political UGC where it is suspected of being unlabelled branded content. Additionally, TikTok plans to update its GPPPA Election Campaign Fundraising Restriction policy in order to improve operability and enforcement. TikTok will also continue to increase the readiness of human and automated moderation systems for emerging election risks, and continue monitoring for urgent election-related incidents, to ensure that election and political-related content cannot be monetised on the Platform; and
- **Article 35(1)(c), Election misinformation**: TikTok plans to revise its Election Moderation Guidelines to address new and emerging product features, as well as to improve policy enforcement across TikTok's Trust & Safety team structure.

14. CIVIC INTEGRITY: RISKS TO PUBLIC HEALTH FROM MEDICAL MISINFORMATION CONTENT

1. Description of the risk:

a. Description of the risk in Year 1

- TikTok understands the risk to be the actual and foreseeable negative effects arising from the dissemination of verifiably (by a recognised medical authority such as the World Health Organisation) false or misleading Medical Misinformation, such as misleading statements about vaccines, inaccurate medical advice that may cause imminent negative health effects such as discouraging people from getting appropriate medical care for a life-threatening disease, and other misinformation that poses a risk to public health (together **'Medical Misinformation'**);
- This risk may arise from attempts to share or disseminate the following content on or through the Platform, whether as short video, comment, livestream or within their profile information:
    - Content undermining the existence or severity of COVID-19 (e.g. that the COVID-19 pandemic is a hoax/scam/exaggerated);
    - Medical Misinformation regarding transmission and prevention of COVID-19 (e.g., that COVID-19 tests cause adverse effects/illness, or that face masks are harmful or will cause illness);
    - Medical Misinformation regarding vaccines, including COVID-19 vaccines (e.g., that COVID-19 vaccines change people's DNA, RNA, or genetic makeup; or that COVID-19 vaccines will be used for mind control/to track people); and
    - Medical Misinformation related to serious medical conditions/life-threatening diseases, including but not limited to COVID-19, HIV/AIDS, Ebola, strokes, cancer, heart diseases, tuberculosis, diabetes, and zika, and other similar viruses or conditions as they may arise; or

- Other Medical Misinformation regarding holistic/homoeopathic remedies (e.g., that drinking or inhaling cleaning/corrosive substances as a preventative or treatment for any disease or that drinking or eating a herbal remedy can treat cancer or other life threatening illnesses).

- In addition, TikTok notes that Article 35 of the Charter enshrines for everyone the right of access to preventive health care and a high level of human health protection as part of EU policies and activities, which may be undermined by Medical Misinformation.

b. Risk Identification: changes to the risk profile since Year 1

In its Year 1 Risk Assessment, TikTok classified Medical Misinformation as a **Tier 2 priority** in its priority tiering system.

This section summarises the risks identified in TikTok's Year 1 Risk Assessment Report and outlines additional risks identified in Year 2. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk detection processes, and extensive consultations with external stakeholders.

Summary of risks identified in Year 1

TikTok identified the following risks as part of its Year 1 Risk Assessment:

- **The risk that content undermines the existence or severity of COVID-19** (e.g. that the COVID-19 pandemic is a hoax/scam/exaggerated);
- **The risk of Medical Misinformation** regarding transmission and prevention of COVID-19 (e.g., that COVID-19 tests cause adverse effects/illness, or that face masks are harmful or will cause illness);
- **The risk of Medical Misinformation regarding vaccines**, including COVID-19 vaccines (e.g., that COVID-19 vaccines change people's DNA, RNA, or genetic makeup; or that COVID-19 vaccines will be used for mind control/to track people);
- **The risk of Medical Misinformation related to serious medical conditions/life-threatening diseases**, including but not limited to COVID-19, HIV/AIDS, Ebola, strokes, cancer, heart diseases, tuberculosis, diabetes, and zika, and other similar viruses or conditions as they may arise; and
- **The risk of other Medical Misinformation regarding holistic/homoeopathic remedies** (e.g., that drinking or inhaling cleaning/corrosive substances as a preventative or treatment for any disease or that drinking or eating a herbal remedy can treat cancer or other life threatening illnesses).

Additional risks identified in Year 2

In Year 2, TikTok identified the following additional risk areas, which it considers to be reflective of the dynamic and evolving nature of Medical Misinformation.

- **Misinformation leading to potentially life-threatening harm**: TikTok has seen an increase in Medical Misinformation claims referring to conditions such as cancer, heart disease, HIV/AIDS, abortion, birth control, pregnancy, serious disabilities, birth defects, strokes and certain chronic

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014662

illnesses such as Alzeimers or diabetes, that could potentially result in serious harm. While this was identified in Year 1, health discussions and interest in topics have shifted away from COVID-19 and include more broadly:

- Misinformation concerning the conditions of a disease;
- Misinformation concerning treatment plans (including dissuading individuals from seeking appropriate treatment and medical care); and
- Misinformation concerning the promotion of unproven remedies.

This type of Medical Misinformation can arise on the Platform, particularly when these topics intersect with other viral trends or developments.

- **Misinformation regarding non-life threatening treatments and alternative prevention/treatment options**: While noting that not all Medical Misinformation will lead to a risk of significant or immediate negative health consequences for individuals and/or society at large, certain claims about health issues and/or treatments can still present a moderate risk and a proportionate approach is still required. Specifically, TikTok has identified the potential for an increase in content which may encourage the use of alternative prevention or treatment methods which are medically unsupported or unproven, or content which misrepresents authoritative sources (i.e. selectively referencing certain scientific data to support a conclusion that is counter to the findings of the study).

c. Inherent Risk in Year 2

For a detailed analysis of how TikTok assessed the baseline severity and probability for Medical Misinformation in Year 1, please refer to TikTok's Year 1 Systemic Risk Assessment Report.

In Year 2, TikTok assesses the overall severity of Medical Misinformation to be 'Moderate' in Year 2. TikTok assesses the probability of Medical Misinformation in Year 2 to be 'Unlikely'. Taking into account severity and probability in Year 2, TikTok has assessed the inherent risk (that is, the risk without any mitigations in place) of Medical Misinformation being shared on the Platform to be 'Low'.

2. Mitigation Measures:

For detailed information on TikTok's system of mitigations/controls (under DSA Art. 35 (1)(a)-(k)) that applied when its designation as a VLOP came into force, please refer to the Year 1 Risk Assessment Report. That Year 1 Report identified further mitigation improvements, progress on which (and any other such mitigations) is listed below. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk monitoring and detection processes, and extensive consultations with external stakeholders.

Implementation of Additional Mitigation Effectiveness Measures in Year 2	
MEASURES IN ACCORDANCE WITH DSA ART. 35(1) A-K	**DESCRIPTION**
(b) Adaptation of terms and	In June 2024, TikTok updated its Community Guidelines to prohibit moderate harm health misinformation, which means false or misleading

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014663

conditions	content regarding the treatment or prevention of injuries, conditions, or illnesses that are not immediate or life-threatening. Alongside this, TikTok updated its moderator guidelines to reflect this change for moderators to align decisions with this new policy.	
(c) Adaptation of content moderation processes	Since Year 1, TikTok has updated its fact checker rating classifications to ensure moderation teams are able to identify Misinformation with more accuracy (read section 13 mitigations). Additionally, TikTok has developed updated moderator guidance to align with updates to the moderate harm health misinformation policy.	
(f) Reinforcing risk detection measures	TikTok has committed to continued risk monitoring for Medical Misinformation as mentioned in Year 1 as a key mitigation improvement. TikTok's primary detection model used for detecting general misinformation, including Medical Misinformation, was updated to factor in the numerous different policy titles from its training data to increase precision.	
(g) Initiating or adjusting cooperation with trusted flaggers in accordance with Article 22 and the implementation of the decisions of out-of-court dispute settlement bodies pursuant to Article 21	TikTok remains ready to engage with EU 'trusted flaggers' as they continue to be designated (read section 5) and in the meantime, TikTok continues to receive reports from Community Partners and subject matter experts.	

3. Residual Risk in Year 2:

Following an assessment of the effectiveness, reasonableness and proportionality of TikTok's mitigations relevant to the systemic risk of Medical Misinformation, TikTok has assessed residual risk to be 'Low' in Year 2. TikTok commits to continuing to expand and invest in mitigation measures to identify and mitigate Medical Misinformation on the Platform.

To support TikTok's assessment of the effectiveness of its mitigations against the systemic risk of Medical Misinformation, TikTok has considered data on its enforcement of its Community Guidelines in the EU for the period of Q3 2023 to Q2 2024, and additional data related to user interactions with video notice tags in order.

Under the Community Guidelines policy of 'Misinformation', TikTok removed 517,353 total videos in the EU, with 502,285 detected and removed proactively, and 411,072 removed without any views.[26] This

[26] TikTok's 'Misinformation' policy prohibits various types of misinformation, including health, climate change, conspiracy and public safety related misinformation. This data has been used for the purposes of both the Medical Misinformation and Harmful Misinformation risk sections. Election Misinformation is actioned separately under TikTok's 'Civic and election integrity' policy.

high proportion of proactive removal suggests that TikTok's content moderation systems remain effective in detecting and removing Medical Misinformation on the Platform, helping to mitigate the systemic risk.

This assessment of the effectiveness of TikTok's proactive detection systems is further informed by the following data on user reports, which serve as a useful indication of the volume of Medical Misinformation that is not proactively detected by TikTok's proactive systems. User reports accounted for a minority of Medical Misinformation detected and removed by TikTok, at 15,068 of the 517,353 total videos detected and removed under the 'Misinformation' policy. TikTok remains committed to continuously working to improve and iterate its proactive detection capabilities, while noting that Medical Misinformation is a rapidly evolving risk area, with nuance across regions and languages.

TikTok's user reporting tools are an important component of its content moderation process, ensuring that Medical Misinformation is detected and removed. The data demonstrates that when TikTok receives user reports of violative content, it is actioned efficiently. Under the 'Misinformation' policy, 9,183 of the 15,068 videos removed following user reports were removed within two hours of receiving the report. Collectively, TikTok's proactive detection and reporting tools have worked effectively to ensure that 406,980 of the 517,353 total videos removed under the 'Misinformation' policy were removed within 24 hours of upload.

The effectiveness, reasonableness and proportionality of TikTok's moderation of Medical Misinformation is indicated by its appeals data, as users can submit appeals against content removals if they believe TikTok has made a mistake. Under the 'Misinformation' policy, only 52,208 of the 517,353 total videos removed were successfully appealed and reversed.

Beyond content removals, TikTok also uses video notice tags to connect users with authoritative sources of information on issues such as Covid-19 and vaccines. Quantitative data reflects strong reach for these notices. Covid-19 notices had 1,894,574,236[27] views, notices related specifically to the Covid-19 vaccine received 985,920,146[28] views, and monkeypox tags accrued 10,156,710[29] total views.

4. Key stakeholder engagement:

For detailed information on how TikTok has engaged with external stakeholders to inform its approach to risk identification and risk mitigation, please refer to Annex 4. This Annex provides an in-depth overview of the collaborative efforts and consultations that have shaped TikTok's strategies in these areas.

5. Prioritisation:

TikTok has reported above on its progress in the last year, and below it states key additional actions for the year ahead to mitigate the risk. TikTok continues to be vigilant to existing risks identified in last year's assessment, including risks identified around medical misinformation that poses significant harm to

[27] Data updated 2 October 2024
[28] Data updated 2 October 2024
[29] Data updated 2 October 2024

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014665

individuals including but not limited to Covid-19 misinformation and vaccine misinformation and has identified additional emerging areas of risk, albeit the overall likelihood of harm has assessed to be comparatively lower to the TikTok community as the risk environment is less concentrated following the end of the pandemic. TikTok has closely considered its risk environment and the inherent and residual risk discussed above and has amended this risk to a Tier 3 risk in this Year 2 prioritisation. However, as noted in Year 1, TikTok will continue to keep its prioritisation under review.

6. Planned further mitigation effectiveness improvements:

- **Article 35(1)(c), Content moderation update:** Over the next year, TikTok plans to expand its work targeting persistent medical misinformation trends by launching a pilot for auto-enforcement in the EU, with appropriate guardrails; and
- **Article 35(1)(b), Policy development**: TikTok plans to develop a new policy for low harm and high prevalence medical misinformation.

15. CIVIC INTEGRITY: RISKS TO PUBLIC SECURITY FROM HARMFUL MISINFORMATION CONTENT

1. Description of the risk:

a. Description of the risk in Year 1

- TikTok understands the risk to be the actual or foreseeable risks to public safety or security arising out of Harmful Misinformation/content which may relate to: armed conflicts and emerging conflicts; acts of terrorism; natural and manmade disasters (such as floods, earthquakes, hurricanes, fires, landslides, environmental or industrial accidents); and other emergency situations that may induce panic, including in relation to current/unfolding events, such as civil unrest (such as protests or riots);
- This risk may arise from attempts to share or disseminate the following content on or through the Platform, whether as short video, comment, livestream or within their profile content that includes:
    - Misinformation making verifiably false and harmful claims regarding natural and manmade disasters (such as floods, earthquakes, hurricanes, fires, landslides, environmental or industrial accidents);
    - Misinformation making verifiably false and harmful claims regarding unfolding shooting events and mass murders;
    - Misinformation making verifiably false and harmful claims regarding public demonstrations or protests;
    - Repurposing old video content, making verifiably false and harmful claims that the event or video is new/current and likely to trigger societal panic (e.g., misleadingly repurposing footage of a bombing or armed attack out of context);
    - Making verifiably false and harmful claims that basic necessities (e.g., food, water.) or services (e.g., banks, cash machines) are no longer available in a particular location, causing hoarding;

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014666

- Stating dangerous conspiracy theories that are violent or hateful, such as making a violent call to action, having links to previous violence, denying well-documented violent events, and causing prejudice towards a group with a protected attribute; or
- Incitement to violence and criminal acts, such as property damage.

- In addition, TikTok notes that Article 6 of the Charter enshrines for everyone the right to liberty and security of person, and Article 12 of the Charter enshrines the right to freedom of peaceful assembly and to freedom of association at all (in particular in political, trade union and civic matters), and that protection of these rights in particular could be restricted or jeopardised in the context of Public Security Risks. TikTok also recognises that in the context of Public Security Risks, individuals retain other fundamental rights under the Charter, in particular the rights to freedom of expression and information, and TikTok's freedom to conduct a business, which must be balanced in a proportionate manner in the context of addressing Public Security Risks.

b. Risk Identification: changes to the risk profile since Year 1:

In its Year 1 Risk Assessment, TikTok classified Public Security Risks as a **Tier 2 priority** in its priority tiering system.

This section summarises additional risks identified in Year 2. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk detection processes, and extensive consultations with external stakeholders.

Summary of risks identified in Year 1

- **Sharing and dissemination of Harmful Misinformation**: The risk that TikTok may be used to attempt to share and disseminate Harmful Misinformation, posing threats to public security in the EU (e.g. inciting riots and civic unrest). This may manifest across TikTok's content formats, including short video or photo, LIVE, comment, or profile features;and
- **Synthetic and Manipulated Media**: The risk that this type of content may be shared on the Platform with additional acute risks of deception and harm; for example, by depicting false instances of current/ongoing violence causing off-platform panic. Content using the likeness of real people may be shared to mislead users about real-world events.

Additional risks identified in Year 2

In Year 2, TikTok identified the following additional risk areas, which it considers to be reflective of the dynamic and evolving nature of Harmful Misinformation.

- **AIGC***:* TikTok identifies new manifestations of harmful and misleading AIGC as part of its ongoing risk identification of synthetic and manipulated media. This risk is monitored closely, in the context of high-profile events and emergencies including elections, armed conflicts, and crises. Potential new manifestations of misleading AIGC include instances where content has been created using an AIGC tool and which features:
  - False depictions of public figures being endorsed or attacked by specific groups or individuals;
  - False depictions of public figures admitting to criminal activity or otherwise making damaging admissions;
  - Catastrophic or crisis events taking place that did not actually occur;

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014667

- Depictions of youth in ways that might re-victimise or fetishise them or violate privacy, for example by depicting victims of crimes or tragedies, or sexualised portrayals of young people;
- Content that appears to come from authoritative or official sources such as news outlets/entities; public institutions and peer reviewed research when it does not;
- Portrayals of public figures being supported or endorsed by individuals or groups that did not do so;
- Competing claims about whether content is authentic or unverified AIGC, particularly that falsely depicts private conversations between public figures; and
- Impersonation, privacy and non-consensual depiction of both private and public individuals.

- **Rapidly evolving events:** TikTok closely considered the content risks which arise from unforeseen and unexpected rapidly evolving events. Content about these events becomes outdated quickly and can potentially be unsafe for users who search for live content. TikTok considers an event to be rapidly evolving if it meets all of the following criteria:
  - Details about the event can change so quickly that, in less than 24 hours from being posted, shared content about situational status and recommended guidance could be considered outdated and potentially unsafe for users to follow;
  - It is distinct from misinformation in that the now out-of-date content was not originally misinformation when posted; and
  - It is an event of multinational impact or concern.
- **Climate change misinformation:** As part of its ongoing efforts to identify and mitigate risks, TikTok defines climate change misinformation as content that denies the existence of climate change or the human activities contributing to it, disputes the scientific consensus on climate change, or spreads false claims about climate change solutions. These narratives undermine public understanding and trust in climate science and hinder efforts to address the global climate crisis.

c. Inherent Risk in Year 2

For a detailed analysis of how TikTok assessed the baseline severity and probability for Harmful Misinformation in Year 1, please refer to TikTok's Year 1 Systemic Risk Assessment Report.

In Year 2, TikTok assesses the overall severity of Harmful Misinformation to be 'Moderate' in Year 2. TikTok assesses the probability of Harmful Misinformation in Year 2 to be 'Possible'. Taking into account severity and probability in Year 2, TikTok has assessed the inherent risk (that is, the risk without any mitigations in place) of Harmful Misinformation being shared on the Platform to be 'Moderate'.

2. Mitigation Measures:

For detailed information on TikTok's system of mitigations/controls (under DSA Art. 35 (1)(a)-(k)) that applied when its designation as a VLOP came into force, please refer to the Year 1 Risk Assessment Report. That Year 1 report identified further mitigation improvements, progress on which (and any other such mitigations) is listed below. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk monitoring and detection processes, and extensive consultations with external stakeholders.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014668

Implementation of Additional Mitigation Effectiveness Measures in Year 2	
MEASURES IN ACCORDANCE WITH DSA ART. 35(1) A-K	**DESCRIPTION**
(b) Adaptation of terms and conditions	As part of TikTok's ongoing risk mitigation for the risk of 'repurposed media' content, in Year 2 TikTok updated its 'Repurposed Media' policy now includes the removal of unedited media content that is presented out of context and may mislead a person about a developing topic of public importance, such as showing a crowd at a music concert and suggesting it is a political protest. TikTok added a policy on 'Misrepresented Authoritative Sources'. This policy prohibits content that promotes misleading correlations or conclusions related to authoritative information that is recognised and trusted, such as reports from research institutions. An example would be content that selectively references certain scientific data to support a conclusion that is counter to the findings of the study. TikTok has updated its AIGC and Edited Media Policy to include prohibitions on: depictions of public figures engaging in or admitting to negative behaviours; claims about public figures' personal characteristics of political or public importance; false endorsements or condemnations; information falsely appearing to come from authoritative sources; and misleading depictions of public importance, including fictitious crisis events.
(c) Adaptation of content moderation processes	TikTok updated its task prioritisation feature and fact checker rating classifications to ensure moderation teams are able to identify Misinformation with more accuracy (read section 13 mitigations). Additionally, TikTok has developed updated moderator guidance to align with these updates.
(f) Reinforcing risk detection measures	As mentioned in Year 1 as a key mitigation improvement, TikTok updated its primary risk detection model, used for detection of general misinformation as well as Medical Misinformation, to better detect sub-categories of misinformation (see Section 13 Mitigations)

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014669

	TikTok also developed an improved system for receiving and utilising third-party intelligence from Fact Checking partners, streamlining the review of actionable intelligence relating to on-platform risk.
(i) Awareness-raising measures and adoption of online interface in order to give recipients of the service more information	TikTok created and launched search guides to minimise the discoverability of Harmful Misinformation regarding rapidly evolving events. These search guides appear to users on the Platform in response to searches relating to rapidly evolving events and assist users in finding authoritative and reliable information. This functionality, launched for the Israel-Hamas war and the Crocus Hall attacks (a terrorist attack carried out against civilians at Moscow's Crocus City Hall on 22 March 2024), can be quickly deployed for crisis events that may be particularly susceptible to misinformation risks and to critical information becoming quickly outdated. TikTok collaborated with fact-checking partners and local media literacy bodies to produce in-app media and digital literacy campaigns to help inform and educate users about identifying and reporting Harmful Misinformation. In August 2023, TikTok rolled out an additional 13 new media literacy campaigns in Europe on topics identified as priority areas for increased digital literacy for TikTok users in Europe. During major events, including COP28 and Earth Month, TikTok adopts an online interface in the form of a pop-up at the top of the search bar, to direct users to an in-app page with some of the most relevant content about climate action. For detailed information on how TikTok has engaged with external stakeholders to inform its approach to identifying and mitigating Public Security Risks, please refer to Annex 4.

3. Residual Risk in Year 2:

Following an assessment of the effectiveness, reasonableness and proportionality of TikTok's mitigations relevant to the systemic risk of Harmful Misinformation, TikTok has assessed residual risk to be 'Moderate' in Year 2.

To support TikTok's assessment of the effectiveness of its mitigations against the systemic risk of Harmful Misinformation, TikTok has considered relevant data on its enforcement of its Community Guidelines in the EU from Q3 2023 to Q2 2024, and additional data relating to video notice tags. TikTok commits to continuing to expand and invest in mitigation measures to identify and mitigate Harmful Misinformation on the Platform.

Under the Community Guidelines policy of 'Misinformation', TikTok removed 517,353 total videos in the EU, with 502,285 detected and removed proactively, and 411,072 removed without any views.[30] This

[30] TikTok's 'Misinformation' policy prohibits various types of misinformation, including public safety, conspiracy, health, and climate change related misinformation. This data has been used for the

high proportion of proactive removal suggests that TikTok's content moderation systems remain effective in detecting and removing Harmful Misinformation on the Platform, helping to mitigate the systemic risk.

This assessment of the effectiveness of TikTok's proactive detection systems is further informed by the following data on user reports, which serve as a useful indication of the volume of Harmful Misinformation that is not proactively detected by TikTok's proactive systems. User reports accounted for a minority of Harmful Misinformation detected and removed by TikTok, at 15,068 of the 517,353 total videos detected and removed under the 'Misinformation' policy. TikTok remains committed to continuously working to improve and iterate its proactive detection capabilities, while noting that Harmful Misinformation is a rapidly evolving risk area, with nuance across regions and languages.

TikTok's user reporting tools are an important component of its content moderation process, ensuring that Harmful Misinformation is detected and removed. The data demonstrates that when TikTok receives user reports of violative content, it is actioned efficiently. Under the 'Misinformation' policy, 9,183 of the 15,068 videos removed following user reports were removed within two hours of receiving the report. Collectively, TikTok's proactive detection and reporting tools have worked effectively to ensure that 406,980 of the 517,353 total videos removed under the 'Misinformation' policy were removed within 24 hours of upload.

The effectiveness, reasonableness and proportionality of TikTok's moderation of Harmful Misinformation is indicated by its appeals data, as users can submit appeals against content removals if they believe TikTok has made a mistake. Under the 'Misinformation' policy, only 52,208 of the 517,353 total videos removed were successfully appealed and reversed

Beyond content removals, TikTok also uses video notice tags to connect users with authoritative sources of information. Quantitative data reflects strong reach; Holocaust video notice tags received 360,164,504[31] total views.

4. Key stakeholder engagement:

For detailed information on how TikTok has engaged with external stakeholders to inform its approach to risk identification and risk mitigation, please refer to Annex 4. This Annex provides an in-depth overview of the collaborative efforts and consultations that have shaped TikTok's strategies in these areas.

5. Prioritisation:

TikTok has reported above on its progress in the last year, and below it states key additional actions for the year ahead. TikTok has closely considered its risk environment and the inherent and residual risk discussed above. TikTok plans to devote extra resources to combating Public Security Risks in the year ahead and as a result it remains a Tier 2 priority. As noted in Year 1, TikTok will continue to keep its prioritisation under review.

purposes of both the Harmful Misinformation and previous Medical Misinformation risk sections. Election Misinformation is actioned separately under TikTok's 'Civic and election integrity' policy.

[31] Data updated 2 October 2024

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014671

6. Planned further mitigation effectiveness improvements:

- **Article 35(1)(c), Content moderation tooling:** TikTok plans to use a new detection system to identify and enforce on re-uploaded content, in order to strengthen the enforcement of AIGC policies, including the detection, removal or labelling of manipulated content such as deep fakes;
- **Article 35 (1)(c),** TikTok will develop a seed bank for Harmful Misinformation content and use a similar detection system to identify and enforce on reuploaded content;
- **Article 35(1)(b), Enforcement on misinformation**: TikTok is planning to develop and test its enforcement on misinformation from a model trained on data from an externally-sourced fact checking repository;
- **Article 35(1)(i), Awareness raising:** TikTok is planning on launching media literacy campaigns in nine more countries, meaning that by the end of the year, it will have run localised campaigns in all 27 member states;
- **Article 35(1)(c), Content moderation updates:** TikTok plans to consolidate the 'Dangerous Misinformation' and 'Dangerous Conspiracy Theory' policy titles to improve moderation accuracy. This consolidation addresses overlaps that have caused moderator confusion, enabling more effective content moderation without altering policy coverage; and
- **Article 35(1)(b), Policy development:** TikTok plans to develop a policy approach for low harm and high prevalence misinformation.

16. FUNDAMENTAL RIGHTS: RISKS TO FUNDAMENTAL RIGHTS

1. Description of the risk

a. Description of the risk in Year 1

- TikTok understands the risks to the exercise of 'Fundamental Rights' on its Platform to comprise the rights set out below, as enshrined in the Charter;
- Following consideration of any actual or foreseeable negative effects for the exercise of fundamental rights as protected under the Charter, TikTok has determined the following fundamental rights as most relevant to its Platform: (1) the right to human dignity; (2) the right to non-discrimination; (3) the right to freedom of expression; (4) the right to private and family life; (5) the right to the protection of personal data; and (6) the right to consumer protection; and
- Within the context of this Report, it is not possible to provide a detailed analysis of each Fundamental Right described above (and as noted above, various other fundamental rights are considered in this Report in the context of other risks).

b. Risk Identification: changes to the risk profile since Year 1

In its Year 1 Risk Assessment, TikTok classified risks to Fundamental Rights as a **Tier 3** priority in its priority tiering system.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014672

This section summarises additional risks identified in Year 2. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk detection processes, and extensive consultations with external stakeholders. These additional risks are described below:

Summary of risks identified in Year 1

- **Content risks:** The risk that content uploaded by Tiktok's users may undermine Fundamental Rights by being discriminatory or dehumanising and/or seeking to scam or defraud other users or other violations of rights;
- **Conduct risks:** The risk of users misunderstanding or ignoring behaviour that is prohibited on the Platform set out in its Terms of Service, Community Guidelines, Privacy Policy and Health, Safety and Transparency Centres; and
- **Moderation risks:** The risk that, in the interest of safety, TikTok's content moderation systems and human moderators may over-moderate by mistakenly acting on speech that doesn't violate Community Guidelines or under-moderate by failing to recognise evolving discriminatory language or other violating content.

Additional risks identified in Year 2

In Year 2, TikTok identified the following additional risk areas, which it considers to be reflective of the dynamic and evolving nature of risks to Fundamental Rights:

- **AI models:** There is an inherent risk that features and advancements in content moderation models may be developed in a biassed or discriminatory manner. This risk arises from the potential development of these features and policies using insufficiently inclusive data sets, safeguards, and safety controls;
- **Armed conflicts:** This is a new emerging risk that is related to the rise in conflicts in various parts of the world resulting in risks to users and human rights defenders. The Israel-Hamas war highlighted room for enhancements in how our policies can best uphold human dignity with regard to content depicting victims of conflict or hostages; and
- **Bias:** Manual policy enforcement may be negatively impacted by political bias, in particular in the context of increased political developments and events, such as the numerous elections in 2024. This bias may manifest itself at the enforcement stage, thus undermining freedom of expression and the right to nondiscrimination, especially when there are increased levels of political and contentious content.

c. Inherent Risk in Year 2

For a detailed analysis of how TikTok assessed the baseline severity and probability for risks to Fundamental Rights in Year 1, please refer to TikTok's Year 1 Systemic Risk Assessment Report.

In Year 2, TikTok assesses the overall severity of risks to Fundamental Rights risks to Fundamental to be 'High' in Year 2. TikTok assesses the probability of risks to Fundamental Rights Year 2 to be 'Possible'. Taking into account severity and probability in Year 2, TikTok has assessed the inherent risk (that is, the risk without any mitigations in place) to Fundamental Rights to be 'Material'.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014673

2. Mitigation Measures:

For detailed information on TikTok's system of mitigations/controls (under DSA Art. 35 (1)(a)-(k)) that applied when its designation as a VLOP came into force, please refer to the Year 1 Risk Assessment Report. That Year 1 Report identified further mitigation improvements, progress on which (and any other such mitigations) is listed below. In the spirit of continuous improvement, these areas of focus result from the risk assessment conducted in Year 1, internal risk monitoring and detection processes, and extensive consultations with external stakeholders.

Implementation of Additional Mitigation Effectiveness Measures in Year 2	
MEASURES IN ACCORDANCE WITH DSA ART. 35(1) A-K	**DESCRIPTION**
(a) Adaptation of platform features or design	As mentioned in the Year 1 Report, TikTok operates an Inclusion Advisory Committee. TikTok's Inclusion Advisory Committee is an internal group made up of employees across the globe with experiences of vulnerable and marginalised communities, including those focused on disability, women's empowerment, racial and ethnic identity, LGBTQ+ inclusion, and more. TikTok has worked with the Inclusion Advisory Committee to ensure marginalised and vulnerable groups' perspectives are considered in feature design, leading to projects like understanding LGBTQ+ experiences with filters and effects, and developing an Accessibility Roadmap for more inclusive product features.
(b) Adaptation of terms and conditions	TikTok also made changes to the Community Guidelines by addressing multiple unfolding and ongoing armed conflicts throughout 2023-24 including adding to their violent actors designation lists, clarifying public interest expectations such as for counter speech by providing examples to moderators, updating guidance on dangerous conspiracy theories and bullying. Also, in relation to the potential risk that inadequate mental and physical health support leads to leakage of violative content on TikTok and/or over moderation of content, the following actions were taken: • A well-being assessment of metric demands to surface potential mitigations (e.g., restructuring to allow for well-being time). This assessment was conducted and resulted in the creation of a user guide outlining the importance of wellness in relation to psychological recovery intended to provide all managers, team leaders and content moderators a clear understanding of the available wellness principles; and • Sharing research and best practice with partners and peers and engaging in cross-industry collaboration on solutions. TikTok is a

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

	member of TSPA[32] and currently engages in their cross-industry group on moderator well-being.
(c) Adaptation of content moderation processes	In November 2023, following a series of discussions with the Jewish Creators Roundtable, TikTok formed a taskforce to enhance its efforts in combating antisemitism, islamophobia, homophobia, and other forms of harm toward marginalised communities. The taskforce aimed to improve the Platform's training methods, focusing on updating implicit bias training with input from external partners. They also implemented new reminders about hate policies and introduced e-learning modules to reinforce these updates. These efforts reduced moderator errors and improved accuracy, with over 75% of moderators finding the training highly effective. TikTok strives to embed human rights standards in policy development and enforcement. For example, in Year 2, TikTok has issued moderator guidance to differentiate Hate Speech from advocacy, thereby protecting freedom of expression. TikTok's Platform Fairness team also conducted the following: • Changed the term 'prostitution' to 'sexual services' in Community Guidelines to reflect Human Dignity recommendations and minimise penalisation of sex workers; • Mapped risks related to human dignity and discrimination, identifying markets prone to Anti-LGBTQ+ sentiment, Anti-immigrant sentiment, punishment of content creators, and censorship of dissent, and proposed recommendations on proportionate content moderation responses; and • Addressed potential biases in policy development and moderation through impact assessments to ensure fairness in Community Guidelines enforcement.
(d) Adaptation of algorithmic systems	In May 2024, the Platform Fairness team revised its Fairness Guidelines by providing specific guidance for different model types, like classification, and included a new section on AIGC. Instead of giving broad advice for all models, the new guidelines offer detailed instructions, which are now part of the model review process. If a model is found to be high risk against this review process, these guidelines are used for fairness testing. To revise these guidelines, TikTok looked at images based on gender expression, skin tone, and body size, and checked how these attributes were distributed across different categories. For each test, TikTok generated 50 images covering various areas where fairness issues often come up, such as jobs, wealth, education, crime, beauty, and more. This helps ensure TikTok's models treat people fairly and align internal processes with Community Guidelines.

[32] https://www.tspa.org/about-tspa/supporters/

(f) Reinforcing risk detection measures	TikTok enhanced its implicit bias training for moderators, particularly around antisemitism and anti-Islamophobia and integrated international human rights standards like the International Covenant on Civil and Political Rights and Rabat Plan in policy reviews. The objective was to protect freedom of expression and prevent discrimination.
Other - right to data protection	TikTok facilitates the right of access through the 'Download Your Data' feature, allowing users to request a copy of their personal data, with additional requests handled via a privacy webform. The right of portability is supported by enabling users to transfer their TikTok data to third parties through an API, in compliance with the EU General Data Protection Regulation ('**GDPR**') and the DSA. Users can exercise the right of erasure by deleting their data or requesting deletion via in-app controls and reporting features. For rectification, users can correct their information through in-app controls or request changes from TikTok. The right of objection allows users to contest data processing for public interest or legitimate interests via a webform. Finally, the right of restriction permits users to request data processing limitations under specific conditions, such as accuracy disputes or legal claims.

3. Residual Risk in Year 2:

Following an assessment of the effectiveness, reasonableness and proportionality of TikTok's mitigations relevant to the systemic risk to Fundamental Rights, TikTok has assessed residual risk to be 'Moderate' in Year 2. TikTok commits to continuing to expand and invest in mitigation measures to identify and mitigate risks to Fundamental Rights on the Platform.

This assessment was conducted on the basis of qualitative and quantitative insights, including data on TikTok's enforcement of its Community Guidelines in the EU from Q3 2023 to Q2 2024. Systemic risks to Fundamental Rights are relevant across all risk sections, and data in those respective sections was consulted in assessing the effectiveness of TikTok's mitigations. For example, the removal of content under TikTok's 'Harassment & Bullying' policy, as documented in the Hate Speech section, is relevant to TikTok's mitigation of risks to the fundamental right to Human Dignity.

Additionally, TikTok has considered data from other relevant CGER policy categories for this risk section. First, under the Community Guidelines policy of 'Suicide & Self Harm', TikTok removed 622,924 total videos in the EU, with 600,489 detected and removed proactively, and 521,259 removed without any views. Second, under the 'Human Exploitation' policy, TikTok removed 64,797 total videos in the EU, with 58,227 detected and removed proactively, and 48,376 removed without any views. Third, under the 'Sexual Activity & Services' policy, TikTok removed 4,213,632 total videos in the EU, with 3,496,944 detected and removed proactively, and 1,461,933 removed without any views. Fourth, under the 'Shocking & Graphic Content' policy, TikTok removed 1,717,094 total videos in the EU, with 1,661,537 detected and removed proactively, and 1,398,038 removed without any views. This high proportion of proactive removal suggests that TikTok's content moderation systems remain effective in detecting and removing violative content negatively impacting Fundamental Rights on the Platform, helping to mitigate the systemic risk.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014676

This assessment of the effectiveness of TikTok's proactive detection systems is further informed by the following data on user reports, which serve as a useful indication of the volume of content negatively impacting Fundamental Rights that is not proactively detected and removed by TikTok's proactive systems. User reports accounted for a minority of relevant violative videos detected and removed by TikTok; at 22,435 of the 49,292 total videos detected and removed under 'Suicide & Self Harm', 6,570 of the 64,797 total videos removed under 'Human Exploitation', 716,688 of the 4,213,632 total videos removed under 'Sexual Activity & Services', and 55,557 of the 1,717,094 total videos removed under 'Shocking & Graphic Content'. TikTok remains committed to continuously working to improve and iterate its proactive detection capabilities, while noting that Fundamental Rights is a rapidly evolving risk area, with nuance across regions and languages.

Fundamental rights, particularly Freedom of Expression, may also be negatively impacted by content moderation that is not reasonable and proportionate. TikTok's appeals mechanism allows users to challenge content removal decisions if they believe TikTok has made a mistake. The reasonableness and proportionality of TikTok's content moderation is demonstrated in the very low share of content removal restrictions that result in successful appeals. Under 'Suicide & Self Harm, 57,439 of the 622,924 total videos removed were successfully appealed. Under 'Human Exploitation', 5,493 of the 64,797 total videos removed were successfully appealed. Under 'Sexual Activity & Services', 95,305 of the 4,213,632 total videos removed were successfully appealed. And under 'Shocking & Graphic Content', 147,776 of the 1,717,094 total videos removed were successfully appealed.

4. Key stakeholder engagement:

For detailed information on how TikTok has engaged with external stakeholders to inform its approach to risk identification and risk mitigation, please refer to Annex 4. This Annex provides an in-depth overview of the collaborative efforts and consultations that have shaped TikTok's strategies in these areas.

5. Prioritisation:

TikTok has reported above on its progress in the last year, and below it states key additional actions for the year ahead. TikTok has closely considered its risk environment and the inherent and residual risk discussed above. TikTok plans to devote extra resources to combatting risks to Fundamental Rights in the year ahead and as a result it has been amended to a Tier 2 risk in Year 2. As noted in Year 1, TikTok will continue to keep its prioritisation under review.

6. Planned further mitigation effectiveness improvements:

- **Article 35(1)(c), Anti-bias in content moderation:** TikTok plans to revamp its training materials for moderators to continually ensure moderation decisions are fair and unbiased.The training update will incorporate new deep dive training on: gender and sexuality, age assurance, race and ethnicity, and divisive political and religious content.
- **Article 35(1)(e), Advertising systems:** TikTok plans to incorporate additional reviews on ads selected and presented to ensure ads are fair and unbiased.
- **Article 35(1)(c), Model FairnessReview Tooling:** TikTok is building an automated system for fairness review of model tickets, resulting in a model fairness review inventory.
- **Article 35(1)(f) Monitoring of Human Rights Risks:** TikTok will establish and implement a framework for ongoing, systematic assessment of country human rights risks, including proactive

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014677

review of local regulations that may conflict with international human rights standards or present risks to TikTok users.

ANNEX 1: KEY INFORMATION ABOUT TIKTOK

What is TikTok?

TikTok's mission is to inspire creativity and bring joy. TikTok is available in many countries globally and its global headquarters are in Los Angeles and Singapore, and its offices include New York, London, Dublin, Paris, Berlin, Dubai, Jakarta, Seoul, and Tokyo. TikTok had on average 142 million monthly active users in the EU between August 2023 and January 2024. TikTok is offered primarily as a mobile app, web app, mobile browser and web browser.

What are the main features of TikTok?

TikTok's video content gives users quick and engaging content experiences whenever they want it. Content is served based on interests and user engagement so entertainment is always personal and connects people from all around the world through shared humour, interests and passions. A non-personalised feed is also available to users in the EU. TikTok's features include video, photo, livestream and comments. Users can share, skip, swipe, like, comment on, or replay videos. Users can also Duet (side by side) with the video of another creator or Stitch another creator's content into their video. Users can send virtual gifts to creators whose content they like.

How does the For You page feed work?

The For You Feed ('**FYF**') is a unique TikTok feature that uses a personalised recommendation system to allow each community member the ability to discover a breadth of content, creators, and topics. In determining what gets recommended, the system takes into account factors such as likes, shares, comments, searches, diversity of content, and popular videos.

TikTok maintains content Eligibility Standards for the FYF that prioritise safety and are informed by the diversity of TikTok's community and cultural norms. TikTok makes certain content ineligible for the FYF as it may not be appropriate for a broad audience and may also make some of this content harder to search for. In order to prevent harmful content from appearing in the FYF, TikTok operates a range of processes and provides an additional layer of controls for minors. The FYF is designed to ensure that users are given opportunities to discover new interests and are not presented with types of content that can create harm when encountered in heavy concentrations.

What does advertising look like on TikTok?

Businesses show ads on TikTok to reach the people they care about in a creative and meaningful way. This helps keep TikTok free for users. TikTok publishes a Guide to Ads and Your Data and is committed

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014678

to being transparent with its users about how it collects, uses and shares data for ads. In Europe, minors do not see ads based on profiling but will see generic ads.

TikTok's advertising policies determine the type of products and services that can be advertised on TikTok. Users will see different kinds of ads when they use TikTok. Users can interact with the ad in much the same way as content posted by other users. For example, users can share, skip, swipe, like, or replay an ad. Users can also comment on an ad if the advertiser enables that feature for a particular ad. If users consider an ad to be violative of TikTok's advertising policies, they can report it to TikTok.

How does TikTok identify and take action on violative content?

TikTok operates proactive and systematic measures to identify, remove or restrict access to content and accounts that violate its Community Guidelines, Terms of Service or advertising policies. TikTok's content moderation is fundamental to its overarching risk management strategy as it underpins its ability to respond effectively to existing and emerging risks. TikTok's risk management strategy places considerable emphasis on proactive content moderation where it endeavours to detect and remove violative content before it is reported by users or third parties.

TikTok operates its content moderation processes using automated and manual (human) means in accordance with the following four key principles, which provide that TikTok will:

1. Remove violative content from the Platform that breaks its rules (whilst noting that TikTok does not allow several types of mature content themes, notably nudity and sexual activity which includes, but is not limited to, pornography);
2. Age-restrict mature content (that does not violate its Community Guidelines but which contains mature themes) so that it is only viewed by adults (18 years and older);
3. Maintain FYF eligibility standards to help ensure that any content promoted by its recommendation system is appropriate for a broad audience; and
4. Empower its community with information, tools, and resources.

TikTok has implemented automated and manual content moderation systems and processes as well as a range of other safety features that are developed, maintained and applied by a range of teams. All video, photo and text-based content uploaded to the Platform are subject to a real time, technology-based automated review. While a video is undergoing this review, it is visible only to the uploading user/creator. Users can report user-generated and ad content which they consider to be violative of the Community Guidelines or TikTok's advertising policies (as applicable).

TikTok's Trust and Safety teams action reports of violative content made by users, non-users and by third parties who form part of TikTok's Community Partner Channel (similar to trusted flaggers under DSA). These teams perform a manual review against TikTok's Moderation Policy Framework, which provides moderators with necessary detail on how to apply TikTok's Community Guidelines. Additionally, in compliance with Article 16 DSA, users can report user generated and ad content that they consider to be illegal under European or Member State law.

TikTok takes action in relation to content that it considers violative, which may include a review by its Trust and Safety team to determine whether the content should be removed or made ineligible for the FYF according to the Community Guidelines. Decisions can be appealed. TikTok also operates a 'strikes' policy for accounts that repeatedly post violative content, which could result in various account level

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

actions up to, and including, a permanent account ban. TikTok adopts a range of processes to review the accuracy of decisions when moderating potentially violative content, in order to ensure that content moderation measures are accurate, effective and proportionate (and in particular to ensure they do not disproportionately impact on users' rights to freedom of expression and information). TikTok's policies and processes for detecting and removing violative ads are similar to those set out above.

An overview of TikTok's data related practices

TikTok offers a suite of privacy settings and controls for users of the Platform. In addition, it has adopted a variety of technical, contractual and organisational measures. TikTok endeavours to ensure the integrity, confidentiality and security of user personal data and to prevent unauthorised access or disclosure of personal data. For example, TikTok implements a multi-tier system of technical access controls (such as system entry controls and technical access controls). TikTok also has in place an incident management program, and forensic capabilities to ensure recovery in case of an incident and the timely restoration of data while ensuring redundancy of the TikTok infrastructure through a backup methodology. Operation and network security are also ensured through the adoption of controls aligned to industry standards, including vulnerability scanning and network monitoring. TikTok has contractual arrangements, with additional supplementary measures in place with group entities and its external service providers to ensure GDPR compliance. TikTok has implemented a number of internal organisational and policy measures to further control access and use of personal data.

TikTok's Privacy Policy applicable to European users of the Platform (and which more broadly covers the European Economic Area, the United Kingdom and Switzerland) provides information and transparency regarding TikTok's processing activities, security measures, and data retention. It also explains to users their rights regarding their personal data, including information on their rights of access, objection, restriction, deletion, rectification and portability of their personal data. This Privacy Policy was updated in November 2023 to address DSA requirements related to sharing data with researchers. Other changes included additional detail for users on the scope and consequences of having a public account and additional information regarding the measures in use to enforce our terms, guidelines and policies.

ANNEX 2: HOW TO USE THIS REPORT

The geographic scope of this Report:

- This Report has been prepared in compliance with Articles 34, 35 and 42 DSA and specifically in accordance with Article 34(2), which requires TikTok to consider the regional and linguistic aspects of any systemic risk in Europe;
- Therefore, the contents of this Report should not be relied upon as representative of TikTok's position outside Europe.

The legal purpose of this Report:

- This Report has been prepared for the limited and specific purposes of Articles 34, 35 and 42

DSA. As such, this Report is not intended to be a definitive statement of TikTok's position on the matters covered, as they may relate to other laws and regulations in Europe;
- This Report should not therefore be relied upon for any other regulatory or litigation purpose, whether inside or outside Europe.

The contents of this Report:

- This Report covers TikTok's risk assessments which were completed by 28 August 2024 and which rely on information collated prior to that date;
- This Report summarises the results of TikTok's detailed risk assessments. It is not intended, and nor should it be treated as, a comprehensive or exhaustive overview of the detailed analysis undertaken in those underlying risk assessments.

Further resources:

- For further information on how TikTok complies with the DSA, please see its European Online Safety Hub (available at: https://www.tiktok.com/euonlinesafety/en/);
- For further information about TikTok's voluntary transparency reporting which shows it enforces its Community Guidelines, please see TikTok's online Transparency Centre (available at: https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2024-1/).

ANNEX 3: SPECIALIST TEAMS

TikTok has a diverse network of internal teams, all of whom have a part to play in its strategy to risk management. The teams below may fulfil their functions in supporting TikTok with its response to one or more Risk Module or systemic risk, as follows:

Trust & Safety Risk & Response team: Trust & Safety's Risk & Response team involves the following key functions, which are summarised as follows:

- Incident Management ('**IM**'): The IM team is a multidisciplinary team of experts who work together to identify, detect, and mitigate risk in response to escalation scenarios. The IM team provides 24/7 incident management and risk handling coverage and support, which primarily involves removal of violating content from the Platform.
- Law Enforcement Response team: TikTok has a dedicated Law Enforcement Response team (known as '**LERT**') with responsibility for reviewing law enforcement and governmental requests in line with applicable law and our policies.
- Law Enforcement Outreach: To enhance education efforts and establish direct communications channels with law enforcement agencies, TikTok has a dedicated Law Enforcement Outreach team. This team frequently engages with national law enforcement authorities ('**LEAs**') across the EU and with international agencies (such as EUROPOL and INTERPOL, and with trusted law enforcement agencies outside the EU) to educate officials on data request submission processes that comply with applicable law and TikTok's policies. An important aspect of outreach engagement is to understand new and emergent trends that help to inform our risk detection and safeguarding strategies.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014681

- Emergency Response Team: TikTok's Emergency Response Team (known as '**ERT**') provides 24/7 coverage to ensure any escalations involving imminent harm or the risk of death or serious physical injury to a person, can be assessed and sent to law enforcement to prevent that harm, as permitted by applicable law. These requests relate to extreme risk, which may involve imminent terrorist-related risks. The team is comprised of specialists who have considerable investigatory and incident-handling experience within the technology industry, a number having previously worked within law enforcement agencies, including the UK's National Crime Agency.
- Risk Analysis team: The Risk Analysis team is staffed by experienced professionals with backgrounds in cyber intelligence and risk detection. This team's role is to monitor risks both in real time and over a period of time to detect emerging trends and patterns, and their activities include monitoring open-source resources and reporting to cross-functional colleagues on potential emerging risks.
- Child Safety team: TikTok's Child Safety Team (known as '**CST**') operates 24/7 to detect, remove and report suspected child sexual exploitation and abuse material to the National Centre for Missing and Exploited Children (NCMEC) and law enforcement authorities. Reports regarding children assessed to remain at imminent and ongoing risk are escalated via priority channels to NCMEC and authorities. The team is comprised of specialists who have considerable investigatory and child safeguarding experience in the technology industry, having previously worked as prosecutors, within law enforcement agencies and child protective and social services.
- Global IP Ops: The purpose of Global IP Ops is to quickly and effectively address any takedown requests regarding alleged copyright and trademark infringements on UGC and to remove any reported content assessed to be infringing with utmost accuracy, in accordance with our terms and conditions / policies and applicable legal requirements. Our aim is to provide a safe environment for our users, uphold the protection of intellectual property rights and safeguard creativity.

Trust & Safety Product Policy: TikTok's Trust & Safety Product Policy team involves four key functions (Issue, Regional, Feature Policy, and Outreach & Partnerships ('**OPM**'), which are summarised as follows:

- Issue Policy Team: The Trust & Safety Product Policy team has a dedicated Issue Policy team that is responsible for TikTok's safety policies. The Issue Policy team includes various specialist teams including:
    - Youth Safety and Well-being Issue Policy Team: The dedicated Youth Safety and Well-being ('**YSW**') policy team are experts in adolescent development, education, and children's rights who consider how Younger Users may be uniquely affected by content, interactions, and platform design features in ways that are developmentally different from experiences for adults. In connection with TikTok's policies and platform features, the YSW policy team's overarching goal is to best protect young people's unique developmental life stages while accounting for diverse global experiences. The team works closely with the other Product Policy Teams, T&S Product, Legal, and other XFN Teams on youth-specific strategies to address harms captured across all T&S policies and feature designs.
        - Content Classification: The Content Classification Policy Team are experts in content classification and maturity ratings and have deep experience and knowledge of existing rating and classification systems, including in Europe, that exist for television, film and games; emerging research into online media and

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

digital entertainment; and developmental psychology. Many of these experts have previous experience rating content at companies like YouTube and Netflix.

- Integrity & Authenticity: The Integrity & Authenticity Team has subject matter expertise on Integrity & Authenticity issues and specific expertise on Election Misinformation issues. Various team members have also undertaken specialist study and research, and have deep sectoral experience in policy development and implementation in this complex and fast-evolving space. The Integrity & Authenticity Issue Policy team leads in the formulation of our policies to combat Election Misinformation.
- Violence & Aggression: Within the Issue Policy team that covers Violence & Aggression, we have sub-teams focused on each of:
    - Exploitation and Abuse: Within this team we have subject matter experts on Exploitation and Abuse who have undertaken specialist study and research, and who have deep sectoral experience in policy development in this field. Together with the Harassment and Hateful Behaviour Issue Policy team, the Exploitation and Abuse Issue Policy team leads in the formulation of our policies that combat harmful content pertaining to GBV Content and plays a crucial role in framing our approach to combating these issues.
    - Violent Behaviours & Dangerous Actors: The Violent Behaviours & Dangerous Actors team has subject matter experts who have undertaken specialist study and research, and who have deep sectoral experience in policy development in this field. The Violent Behaviours & Dangerous Actors team leads in the formulation of our policies that combat harmful content pertaining to Terrorist Content and plays a crucial role in framing our approach to combating these issues.
    - Harassment & Hateful Behaviour Issue Policy: Similarly, the Harassment & Hateful Behaviour Issue Policy team aims to ensure our policies minimise verbal animus and anger directed toward others, in particular marginalised communities. The Harassment & Hateful Behaviour team works closely with the Violent Behaviours & Dangerous Actors team, who together play a crucial role in framing our approach to combating Hate Speech.
- Regional/Local Policy teams: Within our wider Trust & Safety Product Policy team, we also have Regional Policy teams in each region who represent a specific country or region. The Regional Policy teams provide invaluable insight, cultural context, local knowledge and understanding of the interplay between global and local policies, and drive local discussions and engagement on content policy issues. The Regional Policy teams work closely with the Issue Policy teams including, for example, to share local/regional information related to violent extremism-related incidents or threats, hate-speech related incidents or in their local markets, provide policy guidance to moderation teams and ensure that violating content after an incident does not spike in their relevant markets.
- Feature Policy teams: The Feature Policy teams are responsible for developing moderation policies that are specific to the individual products or features that form part of the Platform (such as Comments), as well overseeing account-level policy

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014683

moderation issues. Feature Policy teams work closely with both the Issue Policy teams and Regional/Local Policy teams.

- Trust & Safety Outreach and Partnerships Management team ('**OPM**'): The OPM team is responsible for managing our relationships with existing partners and outreach initiatives and for developing new outreach partnerships. This work is extremely important to our 'Think Global - Lean Local' strategy, as it enables our Trust & Safety teams to access external expertise and build partnerships at a regional and global level, which facilitates those teams in developing policies and practices that are objective, principled and operable. OPM leads several key partnerships and strategic collaboration with external experts, specialists and academics in the field of countering violent extremism, hate speech, minor safety, etc.

Trust & Safety Product:

- Trust & Safety Product: TikTok's Trust & Safety Product team has deep expertise and practical operational experience in designing and implementing safety programmes. The team has deep sectoral experience in policy implementation, threat identification, user safety and crisis handling, and works closely with our Product Policy teams. This level of experience and expertise, and close collaboration with the Product Policy teams is crucial in the formulation of effective strategies to detect and combat systemic risks. This team leads on the detection of and enforcement against harmful content and the implementation of TikTok's safety-by-design strategy. Within the T&S Product Team is a dedicated function for product implementation of Age Assurance measures. The Age Assurance team's efforts are focused on ensuring TikTok has greater confidence in user age for higher risk products and experiences, enabling us to build age-appropriate content and feature experiences.
- Covert Influence Operations team: The Covert Influence Operations Team is a key team working to disrupt external attempts to engage in covert influence operations by manipulating the TikTok Platform and/or harmfully misleading our community.
- LIVE Safety: TikTok's LIVE Safety team also comprises product managers, algorithm engineers, engineers and data scientists working in close collaboration, along with the LIVE Policy team, to design safety strategies, as well as launch features and models. We also monitor for new livestream risk trends and/or potential viral harmful content via media monitoring and risk monitoring mechanisms.
- Regional Safety & Integrity team: Within our Trust & Safety team, we have Regional Safety & Integrity teams. Led from Dublin, the role of the EMEA Regional team is to monitor and manage risks at a regional level, including across the EU, and to design, coordinate and project manage cross-functional risk mitigation programmes, involving multiple other teams (e.g. such as programmes focused on proactively managing, scenario-planning and mitigating risks around large events or major sports tournaments, events/dates/anniversaries). The EMEA Regional Safety & Integrity team plays an important role in risk detection and identification and longer-term mitigation at a regional and local level through its proactive scenario-planning, programme management and wider cross functional collaboration.
- Trust & Safety Operations (content moderation teams): An important part of content moderation involves human review by our trained content moderation teams. No matter the time of day, if content is reported or flagged for review, our teams are on standby to help take appropriate action. To ensure 24/7 coverage, we operate a 'follow-the-sun' model. Through human review, we also improve our machine learning systems as moderators provide feedback to the

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014684

technology, helping to capture emerging content trends and improve our future detection capabilities.

- Account moderation teams: While the technical measures described above in this section are a core aspect of our approach on the Platform, an important part of moderation involves human review by our trained content moderation teams. Suspected underage accounts, including those reported via the in-app reporting channel, are reviewed by the moderation teams handling account-level moderation.
- Product Team: Sitting within the Product organisation, the Privacy & Responsibility Team concentrates on: minor protection and age-appropriate access; data transparency, management and access; and user choices over personalisation, interaction, and discoverability.
- Global Business Solutions - CoreOps: The CoreOps team conducts further investigations into advertisements, as required, to investigate hidden risks. The CoreOps team may action the advertiser account depending on the severity of the violation found.
- Global Privacy Operations: This multilingual team, which is led from Dublin, provides support to our users for any privacy-related inquiries including requests to exercise the GDPR rights of access, deletion, restriction etc.
- Public Policy and Government Relations: TikTok's regional and local Public Policy and Government Relations teams play an important role in bringing local and regional context to our decision-making, including feeding input from regulators, government departments and other institutions and external stakeholders to Trust & Safety, Product and other parts of TikTok. Our work in this area is also supported by TikTok's Public Policy Subject Matter Experts team, which has deep expertise at the intersection of public policy, content regulation and harmful content issues, and who provide support on the public policy aspects of TikTok's work in this space.

ANNEX 4: STAKEHOLDER ENGAGEMENT

1. INTRODUCTION 94

2. STAKEHOLDER ENGAGEMENT: Purpose and Role 95

3. ILLEGAL CONTENT: Risks of Illegal Hate Speech Content 96

4. ILLEGAL CONTENT: Risks of Terrorist Content 98

5. ILLEGAL CONTENT: Risks of Intellectual Property Infringing Content 99

6. ILLEGAL CONTENT: Risks of Child Sexual Abuse Material and Child Sexual Exploitation and Abuse 100

7. ILLEGAL CONTENT: Risks of Gender Based Violence Content 101

8. YOUTH SAFETY: Risks Related to Age Appropriate Content and Online Engagement 103

9. YOUTH SAFETY: Risks related to Age Assurance 105

10. CIVIC INTEGRITY: Risks to Elections and Civil Integrity 107

11. CIVIC INTEGRITY: Risks to Public Health from Medical Misinformation 108

12. CIVIC INTEGRITY: Risks to Public Security from Harmful Misinformation Content 109

13. FUNDAMENTAL RIGHTS: Risks to Fundamental Rights 110

14. Generative AI 111

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014686

1. INTRODUCTION

TikTok recognises that the complex nature of potential risks arising from platforms hosting and sharing UGC requires constructive engagement with a wide range of outside experts and stakeholders. Seeking advice and input across a wide array of stakeholders has been a cornerstone of the development of TikTok's safety efforts prior to the DSA and continues to be a central pillar of its risk governance efforts. These engagements are integral to TikTok's process of developing and refining its safety policies and operations and continuously improving the Platform's capacity to reduce and mitigate the risk of harm. TikTok's stakeholder engagement strategy reflects a commitment to dynamically adapting its services based on continuous feedback and collaboration.

Over the past year, TikTok has been consistently proactive in engaging with stakeholders across various regions and contexts to ensure that it has a comprehensive understanding of the diverse challenges and opportunities that arise from its nature, scale and impact. In alignment with Articles 34, 35, and Recital 90 of the DSA, TikTok has - and will continue to - prioritise external stakeholder engagement in its risk assessment and mitigation strategies.

This Report details TikTok's strategy to engage with stakeholders to ensure that measures to mitigate systemic risks on the Platform are reasonable, proportionate, and effective, whilst also reinforcing users' right to freedom of expression. This strategy involves maintaining ongoing dialogues with users, policy makers, industry experts and civil society organisations. TikTok's own ESAC is also key as that body provides critical expert input into both identifying risks specific to TikTok and in advising upon specific mitigation implementation.

This stakeholder engagement report covers the following topics:

- The relevance of stakeholder engagements in effectively identifying and mitigating risks on TikTok;
- An analysis of the most relevant stakeholder engagements conducted in Year 2 of the DSA, organised by TikTok's understanding of the systemic risks outlined by the DSA; and
- Insights from consultations with ESAC members regarding TikTok's approach to DSA systemic risk assessments.

2. STAKEHOLDER ENGAGEMENT: PURPOSE AND ROLE

Developing Trust and Safety policies and processes to support the many communities that are celebrated on TikTok is not something it does alone, as TikTok purposely seeks to engage with as many different perspectives as possible at all steps of the development process.

This dialogue is important for a number of reasons. While TikTok has an experienced and knowledgeable Trust & Safety team, external perspectives are essential to help it more deeply understand and proactively address topics of concern. TikTok is committed to listening to and including independent voices and to ensure that it makes decisions rooted in fact and aligned with international

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014687

best practices. These partnerships come from a range of entities, such as NGOs, industry experts, academia, researchers, creators and the voice of lived experience. TikTok manages these engagements in a variety of ways, including ESAC and a separate Youth Council (which launched on 25 March 2024), one off consultations, long term partnerships, roundtable discussions, focus groups and community engagements. Diversifying the ways and means in which TikTok gathers these external inputs allows it to support its Trust and Safety teams in a range of different ways, appropriate to policy areas of focus,

Third party engagement is also important in building trust across stakeholders and importantly with TikTok's user community. With expert input, TikTok can transparently demonstrate its efforts to understand the topic in question and to develop a solution that is balanced and factual. This process provides reassurance in the content decisions that TikTok makes and provides additional safeguards that prevent harm before it happens, improving overall user experience. TikTok also provides post-implementation feedback to its external partners where possible so that those partners can understand and observe the impact of their engagement.

3. ILLEGAL CONTENT: RISKS OF ILLEGAL HATE SPEECH CONTENT

TikTok is committed to ensuring the safety and well-being of its users by undertaking stakeholder engagements to understand and mitigate Hate Speech.

a. **Marginalising speech and behaviour**

Between December 2023 and April 2024, TikTok conducted nine consultations on marginalising speech and behaviour with 35 experts from 14 countries. As part of this work, TikTok engaged with at least one member of every TikTok content advisory council across the globe including **ESAC**. TikTok also spoke with community organisations and content creators affected by Hate Speech on the Platform, as well as behavioural scientists, and freedom of expression experts. Based on the concerns raised by the external experts around disproportionate restriction of speech, TikTok re-defined and limited the scope of its Marginalising Speech and Behaviour policy. The policy drafters clarified what type of content the policy should address, and provided more clear and detailed instructions on how to categorise content under this policy. The risk of over enforcement may create risks of harm to freedom of expression, which this work sought to limit. The team defined and limited the use case scenarios for the policy to only periods with a high risk of harm and increased Hate Speech and is developing extra guardrails for the implementation of the policy to assure consistent and reliable decision-making. The policy launched in May 2024. TikTok subsequently updated its Community Guidelines, which now states that 'Some content that uses stereotypes, insinuation, or indirect statements that may implicitly demean protected groups' is ineligible for the FYF.

b. **Hate Speech mitigation**

TikTok conducted a dedicated consultation in June 2024 to seek advice and feedback on its risk assessment and management approach with a specific focus on TikTok's current approach to identifying and mitigating Hate Speech. TikTok consulted with Maja Heban, a member of ESAC and a LGBTQIA+ rights activist. Heban validated TikTok's effectives use of keyword filter tools that empower users to block hateful comments on content and livestreams proactively.

TikTok has continued to participate as a signatory to the EU Code of Conduct on Countering Illegal Hate Speech Online. In collaboration with the European Commission and other participating platforms, it has been actively involved in the revision and improvement of the Hate Speech Code since 2020. TikTok has committed to sign up to the revised Hate Speech Code, which is intended to be adopted under Article 45 DSA later this year.

TikTok undergoes yearly 'Hate Speech test' evaluations by participating EU based NGOs, who conduct Hate Speech monitoring over a four-to-six week period to assess the Platform's response time and reliability in addressing reported content. While the Code is being finalised for DSA adoption and formal monitoring is paused, TikTok has participated in the SafeNet programme run by INACH, an independent platform monitoring initiative. TikTok regularly engages with INACH and the participating organisations on this initiative to receive their feedback, until formal monitoring commences again.

As part of the new revised Hate Speech Code, TikTok has committed to significant new obligations. These include a commitment to review 50% of valid illegal Hate Speech user notices within 24 hours, demonstrating TikTok's dedication to swiftly addressing illegal content on the Platform.

TikTok also held roundtables with Muslim, Jewish, and LGBTQ creators to understand their experiences and address toxicity on the Platform. TikTok's Law Enforcement Outreach team engaged with the European Counter Terrorism Centre and the EU Internet Referral Unit to address potential Terrorist Content, and collaborated with the Hate Speech Investigation Group in Rheinland Pfalz, Germany.

c. **Antisemitism and Islamophobia emerging from the Israel-Hamas war**

In order to validate its approach to enforcing Hate Speech policies within the context of the Israel-Hamas war, TikTok engaged with 14 experts from Jewish, Arab, and Muslim communities, as well as freedom of expression and human rights specialists. This included consulting Shannon Raj Singh, an expert on digital technology and mass violence. TikTok used this expert feedback to inform its enforcement of its Hate Speech policies related to the conflict ensuring it is as balanced as possible, particularly with regards to religious expression.

d. **Protections of Public Figures**

TikTok engaged with over 40 experts, including members of ESAC, to review its definition of public figures and their level of protection under TikTok's Bullying & Harassment Policies. Like other platforms, TikTok strives to strike a balance between allowing freedom of expression - and legitimate criticism of public figures - without neglecting its responsibilities to protect these public figures from abuse. TikTok defines public figures as: 'Adults (18 and older) with a significant public role, such as a government official, politician, business leader, or celebrity.' TikTok does not consider people under the age of 18 as public figures. This approach was accepted and endorsed by the experts consulted. The majority of experts supported distinguishing between public and private figures, but the extent and method were debated. Proponents of a higher bar for criticising public figures argued that such individuals, by their role, invite scrutiny, and that limiting criticism could harm democratic debate. Supporters of greater protections were concerned about the chilling effect on women and vulnerable communities, the mental health of public figures, and the potential for toxic discourse on online platforms. Even among those who disagreed on the level of protection for public figures, there was consensus that the definition of a public figure could be refined, with a majority agreeing on increased

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014689

protections in certain areas, such as body shaming. There was broad agreement that minors should not be considered public figures. This review provided valuable insights and highlighted diverse perspectives on the issue. This work has validated our broad approach to protections of public figures under our Harassment and Bullying policies, and is continuing to inform the planned evolution of this policy with potential tweaks to the policy possible later in 2024.

e. **Reporting Hate Speech**

TikTok has over 30 EU-based organisations to monitor and report suspected Hate Speech as part of its Community Partner Channel. TikTok is aware that Hate Speech is a cross-platform issue, as actors and content can spread between platforms, leading to a cumulative incident which one platform would not be able to identify on its own.

f. **Campaigns to raise awareness**

TikTok signed a new partnership with the Six Nations rugby tournament to continue to run Swipe Out Hate campaigns including for the 2024 editions of the tournament. TikTok ran these campaigns to educate TikTok's users on its zero tolerance approach to hate and encourage its users to report hateful content during both the men's and women's Six Nations tournaments.

4. ILLEGAL CONTENT: RISKS OF TERRORIST CONTENT

The moderation of terrorist and violent extremist activity and content online is an especially nuanced and dynamic area which requires deep understanding and linguistic knowledge of ideologies, political speech, and violent extremist use of technologies. As such, TikTok engages with a variety of subject-matter experts to develop proactive and reactive policies to address the risk of Terrorist Content on its Platform.

a. **High-risk events and crisis readiness in the EU**

TikTok has actively engaged in several strategic initiatives in order to enhance its response to EU terrorist attacks and crises. In February 2024, TikTok participated in a Terrorist Content Online Table Top exercise with the EU Internet Referral Unit (EU IRU), Europol, member states' law enforcement representatives, and observers from the UK and New Zealand. This exercise tested readiness for large-scale terrorist attacks, focusing on content moderation and proactive escalations.

In September 2023, during Referral Action Day, TikTok supported Europol's European Counter Terrorism Centre and the EU IRU, alongside 11 EU Member States, in targeting and assessing 2145 pieces of suspected terrorist and violent extremist content.

TikTok also contributed to the SIRIUS EU Electronic Evidence Situation Report, collaborating with Europol and providing input on improving cooperation between EU Member States and social media platforms. Meetings with the SIRIUS unit, including a session at TikTok's Dublin office in April 2024, reinforced this collaboration. TikTok provided strategic guidance to law enforcement agencies across Europe ahead of the Olympic Games and Euro Football championships, which subsequently facilitated

data requests and emergency response strategies.

TikTok also engaged with counterterrorism and violent extremism agencies from various EU Member States, including France, Germany, Spain, Belgium, Netherlands, Italy, and Ireland. These engagements aimed to educate law enforcement on how TikTok's processes comply with relevant DSA provisions and to establish efficient communication channels. Turning to the risk of a crisis event happening in the context of an election of which there is a significant volume in 2024 - TikTok is a member of the Code of Practice on Disinformation (CoPD) Working Group on Crisis and also co-chairs the Elections subgroup to collaborate with the European Commission and signatories on trends and mitigation strategies. Also, TikTok is an active member and participant of the EU Internet Forum. Through this engagement, TikTok engages with industry and government for dialogue and the exchange of best practices.

b. **Threats related to evolving Terrorist Content**

TikTok has worked extensively with Tech Against Terrorism to continue understanding terrorist and violent extremist behaviour and tactics online, adversarial shifts and mitigation strategies. Since September 2023, TikTok has conducted 7 workshops with Tech Against Terrorism to address:

- Educational, Documentary, Scientific, and Artistic Policies to assess challenges around such as counterspeech content, human rights implications and content moderation evasion techniques;
- 'Borderline content', or content that is not evidently related to Terrorist Content, activity, or groups, or not obviously violative of Community Guidelines but may be shared by a terrorist entity;
- Violent speech against designated terrorist entities; and
- The strengthening of human rights safeguards when identifying and removing Terrorist Content online.

TikTok engaged with NGO Amadeu Antonio Stiftung from February 2024 to address terrorist and violent extremist threats in Germany and Austria. This initiative identified local signals of hateful ideologies and actors in Germany in order to enhance moderation accuracy by providing localised guidance on signals that can be detected in order to identify violations at scale. This collaboration also identified specific audio files and songs as potentially violative of TikTok's policies.

c. **ESAC advice on emerging threats**

ESAC, together with TikTok's Safety Advisory Council members across the Middle East and the Asia Pacific regions, have been continuously consulted in relation to ongoing crises such as the Israel-Hamas war so that TikTok can ensure that its crisis response is as well informed as possible. For example, in December 2023 TikTok convened the first issue-specific Safety Advisory Council meeting (with members from across all regional councils attending) on terrorist entities and activity online.

In June 2024, TikTok conducted a dedicated consultation with TikTok's ESAC member Judy Korn, the founder of the Violence Prevention Network, who has dedicated over 20 years to working with violent extremist offenders and monitoring their online activities. Korn provided critical feedback on assessments of borderline content and conduct-related risks related to known behavioural patterns of terrorist use of UGC platforms.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014691

5. ILLEGAL CONTENT: RISKS OF INTELLECTUAL PROPERTY INFRINGING CONTENT

TikTok collaborates closely with external experts and stakeholders to manage the risks associated with IP-Infringing Content to ensure compliance with EU and national laws. By engaging with these specialists, including content creators and IP-rights holders themselves, TikTok strives to effectively combat the unauthorised use of protected material, such as music, art, audio-visual recordings, trademarks, and patents.

a. **Industry collaborations and regulatory engagements for IP-Infringing Content**

TikTok's Trust & Safety team and Government Relations Public Policy Safety teams actively engage with various third-party industry bodies and external experts to assess and mitigate the risks associated with IP-infringing content. These collaborations provide crucial information and insights that inform TikTok's policies and ensure that its Community Guidelines are both objective and balanced. Key engagements include those with the Audiovisual Anti-Piracy Alliance (AAPA), an industry group consisting of rights owners and broadcasters, which focuses on tackling audiovisual piracy in Europe through lobbying, supporting law enforcement, and building partnerships; and the Union des fabricants (UNifab), a French association dedicated to promoting and defending intellectual property rights.

TikTok cooperated with the Copyright Information and Anti-Piracy Centre (CIAPC), a Finnish non-profit association, which was appointed as a Trusted Flagger under DSA Article 22 by the Finnish Digital Services Coordinator, Traficom on 7 March 2024. Since February 2024, TikTok has processed CIAPC's IP reports with priority, although CIAPC has not submitted any reports under Article 16 DSA to date.

TikTok engages with regulatory authorities within the EU such as France's Arcom and Italy's AGCOM in relation to our IP protection measures (including compliance with the EU Copyright Directive).

6. ILLEGAL CONTENT: RISKS OF CHILD SEXUAL ABUSE MATERIAL AND CHILD SEXUAL EXPLOITATION AND ABUSE

TikTok actively engages with stakeholders focused on controlling CSAM dissemination and preventing CSEA related harms. TikTok maintains close relationships with a number of expert child safety groups to identify potential risks before they reach TikTok as they are detected across other platforms and services.

a. **Key industry partnerships**

TikTok is a member of the WeProtect Global Alliance in order to help build collaboration across platforms. TikTok's active participation in the Alliance fosters collaboration across sectors – including NGOs, safety firms, and governments – facilitating the sharing of knowledge and best practices, and strengthening the collective response to CSEA. TikTok is also a member of the Internet Watch Foundation (IWF) and the Safer Internet Centers. TikTok regularly receives and acts on keyword lists and hash lists from organisations including the National Center for Missing and Exploited Children (NCMEC) and IWF to ensure that internal detection tooling is running on the basis of most up-to-date emerging risk signals.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

b. **Information sharing**

TikTok holds a board seat on the Tech Coalition, an industry group dedicated to fighting CSEA online. TikTok participates regularly in – and leads several working groups – within the Coalition. TikTok has also led work for the Tech Coalition and other tech companies in relation to hackathons, conferences, and other events. The output of such activities serves to increase TikTok's information ability to recognise and remove CSAM more quickly, for example regarding grooming behaviours in regional contexts. With the Tech Coalition, TikTok has contributed to the creation of a risk identification rubric in order to identify potential areas of increased risk of CSAM or CSEA, for example whether a product enables real-time communication or whether it enables text or multimedia based content sharing and interactions.

c. **Identifying emerging risks**

TikTok participates in consultations and meetings with the Institute of Sexology and Sexual Medicine and the UK and Ireland chapter of Stop It Now! to better understand and identify sex offender behaviour and juvenile sex offender psychology to proactively address and prevent instances of CSEA on TikTok.

TikTok has also undertaken consultations with InHOPE on AIGC CSAM and grooming patterns associated with AIGC CSAM. This helped TikTok create appropriate language for its CSAM policy. TikTok also participated in INHOPE's conference, sending staff to present and learn best practices from those managing CSAM hotlines.

TikTok also engages with ConnectSafely for insights and/or recommendations on relevant youth safety topics to help TikTok better understand emerging risks, trends and/or relevant world events that might impact users on its Platform.

d. **Third-party detection tools and keyword lists**

TikTok's Child Safety Team meets monthly with NCMEC to coordinate on best methods for CSAM and child endangerment/abuse response. TikTok receives direct feedback on the efficacy of its reporting, as well as information when reports have led to specific children being identified and helped. TikTok made 288,125 reports to NCMEC in 2022 and in 2023, that number grew to 590,376. TikTok also participates in NCMEC's Take It Down initiative to share known hashed CSAM content to help prevent CSAM from being disseminated on the app. Some of the CSAM Detection Systems that TikTok deploys and implements are integrating PhotoDNA and the Google Content Safety Toolkit, along with ingesting hash lists from the IWF. These enable TikTok to remove previously-identified CSAM at point of upload, before it has been viewed.

7. ILLEGAL CONTENT: RISKS OF GENDER BASED VIOLENCE CONTENT

TikTok is committed to the prevention of Gender Based Violence (**'GBV'**) Content occurring on its Platform. GBV Content disproportionately affects women and marginalised communities, and addressing such hateful content is crucial for promoting gender equality and ensuring the safety and well-being of all individuals. In the past year, TikTok has engaged with expert stakeholders from leading

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014693

nonprofits that are working to prevent GBV and to support survivors. These engagements have supported the identification and categorisation of granular issues under the broader umbrella of GBV Content, supported the roll out of new tools (for example a hash matching system), new resources in the Safety Center, and assisted with raising awareness, especially for Younger Users. Finally, a cross-platform partnership has led to a larger case study on the risk of sextortion.

a. **Sextortion**

Sextortion is a type of blackmail in which the perpetrator has (or purports to have) sexually explicit content about a user and then uses threats (to distribute that content) and intimidation to extort money from that user. Sextortion is a known risk that TikTok prohibits, and one that requires continuous monitoring due to its evolving nature as bad actors seek new strategies. The number of global sextortion cases across the internet and several platforms is increasing rapidly. Reports to the NCMEC more than doubled in 2023 across public and electronic service providers, rising to 26,718 compared to 10,731 the year before. As a result, TikTok partnered with Toluna to carry out a user survey in multiple countries to better understand the scope of the problem on its Platform. A case study on this topic is included in the GBV section of the Report above.

b. **Non-Consensual intimate imagery and image based sexual abuse**

TikTok continues its partnership with StopNCII for proactive detection when an image is uploaded, and so that TikTok can receive and enforce on known examples of non-consensual intimate imagery. TikTok then employs a hash matching system to proactively identify and remove potential non-consensual intimate images on the Platform.

TikTok attended two conferences, the Virtual Summit on Deepfake Abuse and Silicon Saviors or Digital Threats? Exploring Tech's Impact on Domestic and Sexual Violence, which informed its identification of the risk of image based sexual abuse. Image Based Sexual Abuse ('IBSA') is the creation, manufacture, or distribution of nude, partially nude, or sexually explicit content without the consent of the person in the content, for the purpose of sexualizing their body, or portraying them in a sexual manner. TikTok decided to approach the harm of AIGC images the same regardless of whether it includes real or manipulated media, as the impact on the survivor is the same. TikTok updated its policies to make this clear.

c. **Gendered Harassment and Youth**

TikTok co-hosted a roundtable discussion on gendered harassment and the impact of misogynistic content on youth together with **Redacted** from the George Washington University Law School and Cyber Civil Rights Initiative. Representatives from a range of nonprofits attended, including experts from A Call to Men, Futures without Violence, PreventConnect, Men Can Stop Rape, RALIANCE, the National Sexual Violence Resource Center, and the National Domestic Violence Hotline. These representatives work on specific issues within GBV and each brought varied and specialised perspectives to the discussion. Their insights highlighted nuances and granular aspects of this complex issue, which inform TikTok's broader risk identification strategy. Educating youth about respect, consent, and healthy relationships can have a lasting impact, reducing the prevalence of GBV long term. TikTok further discussed youth focused GBV prevention with Futures without Violence. These discussions highlighted nuances relevant to TikTok's policies at the intersection of youth safety and GBV content.

d. **Language promoting GBV**

TikTok held policy consultations on the risks associated with the promotion of GBV content with each of the following experts: **Redacted** from Stop It Now! **Redacted** member of TikTok's Latin

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014694

America Safety Advisory Council, and Redacted CEO/Co-Founder of Glitch and member of TikTok's ESAC. These meetings identified and categorised the harm levels of different forms of language that promote GBV Content and the psychological impact on users viewing such content, which informed TikTok's ongoing policy-drafting process.

e. **Safety Center and Survivor Resources**

TikTok held two meetings with No More to consult with them on the existing resources available to survivors in the Safety Center. A consultation with Redacted Executive Director of the DC Rape Crisis Center and Board Member of the National Alliance to End Sexual Violence revealed that TikTok's Safety Centre pages on sexual abuse could be improved by using more survivor-centred language in order to avoid exacerbating survivors' trauma and hindering their recovery. TikTok partners with local organisations, helplines and social services in individual EU Member States to tailor its in-app support resources for survivors. These include rape crises centers across multiple EU member states.

f. **LGBTQIA+ Community**

TikTok works with Glaad and Stonewall to gain insight into the lived experiences of the LGBT+ community so that it can mitigate the risk of GBV in a socially and culturally sensitive manner. LGBT+ individuals often experience GBV differently, and transgender and non-binary people are at a higher risk of violence. These partners can also report abusive content to TikTok and help to identify new trends. TikTok's Community Partner Network entities that assist with this are Stopline (Austria), BeLong to Youth Services (Ireland), Arcigay (Italy), FELGBT (Spain), Zara (Austria), Stop Fisha (France), and The International Network Against Cyber Hate.

8. YOUTH SAFETY: RISKS RELATED TO AGE APPROPRIATE CONTENT AND ONLINE ENGAGEMENT

TikTok is committed to ensuring the safety and well-being of its Younger Users by actively engaging with expert stakeholders to understand and mitigate risks related to Age Appropriate Content and risks related to Online Engagement. Collaborating with experts has enabled TikTok to identify and mitigate emerging risks, such as detecting mature content and preventing young audiences from accessing inappropriate material. Stakeholder input has been vital in addressing exposure to harmful content, discoverability by bad actors, and mental health impacts from features like comments and subscriptions.

a. **Age Appropriate Content**

TikTok closely observes the work of global content ratings bodies to inform its policies and processes to prevent Younger Users' exposure to content that can be harmful or inappropriate. These bodies are public organisations that classify films, videos, and other online content and motion pictures into age-based ratings. There are several global ratings bodies whose guidance TikTok actively monitors, including but not limited to the BBFC, the Netherlands Institute for the Classification of Audiovisual Media, the Motion Picture Association, the TV Parental Guidelines in the US and the Korea Media Rating Board. By following ratings bodies globally, TikTok identifies when changes to national guidelines are made so that any applicable changes may be reflected in TikTok's Content Classification system.

b. **Online Engagement**

TikTok is part of multistakeholder forums which bring together representatives from countries,

intergovernmental organisations, civil society organisations and the technology industry to identify, implement and share innovative solutions to better protect children online. The consultations and emerging topics and trends from this inform TikTok's identification of risks to minor safety.

c. **Campaigns for Younger Users**

TikTok recognises the importance of reaching Younger Users in ways that resonate with their unique cultural and social circumstances. To achieve this, TikTok partners with local organisations in individual EU countries to tailor its approach.

In Italy, TikTok works with the Safer Internet Centre, which promotes a safer and better use of the internet among children, parents, and teachers. TikTok has worked with e-Enfance in France to combat cyberbullying by integrating the national hotline for digital violence victims directly onto its interface. Centro Internet Segura in Portugal operates the Linha Internet Segura, a telephone and online support service for young people online. It is part of TikTok's Community Partner Channel where onboarded organisations can submit reports for prioritised action.

d. **Resources for parents on youth safety**

TikTok has partnered with the UK-based NGO Internet Matters to develop policies and in-app features across the EU-UK market that seek to improve digital well-being. Through this partnership, TikTok creates resources to educate families about the importance of digital well-being and addresses common challenges such as peer pressure online. In partnership with the Family Online Safety Institute, TikTok created The Good Digital Parenting Initiative, offering resources to help parents better understand the available tools and controls that can shape the desired digital environment for their family. The continued collaboration with the Digital Wellness Lab at Boston Children's Hospital further supports these safety improvements. The Lab has consulted on various platform designs and digital tools. Most importantly, the Lab provided insights that contributed to TikTok's design of its screen time management tools, which imposes a default 60-minute limit for Younger Users and several different options for Family Pairing accounts. TikTok is also an early signatory to the Lab's Inspired Internet Pledge, through which it has committed to focusing on well-being for Young Users, and sharing best practices on creating a healthy online environment.

TikTok also conducted consultations and engagements with 13 organisations worldwide to guide and share developments around family pairing product updates.

e. **AIGC tools**

TikTok partners with the NoFiltr Youth Innovation Council, which is an initiative that brings together youth globally to share their insights and advice to make the internet safer. The Council is a unique space that provides the chance for youth voices to be heard on issues that directly affect them and allows TikTok to engage directly with affected young user groups on various experiences with the app to inform product and policy development.

TikTok's Youth Safety and Well-being Policy team worked in collaboration with the Outreach and Partnerships team to conduct two consultations with youths via NoFiltr in order to identify potential risks related to AIGC tools This engagement has informed TikTok's understanding of the risks to self-esteem, body image, and mental health.

TikTok also consulted with various mental health and safety advisors on the use of beauty filters on the

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Platform. Redacted spokesperson for the Department of Mental Health of the Ministry of Public Health of Thailand, and members of TikTok's Brazil Safety Advisory Committee, Fabiana Vasconcelos and Redacted shared their expertise on the use of photo and video effects.

f. **TikTok's Youth Council and teen well-being**

On March 25, 2024, TikTok launched a global Youth Council to reinforce its commitment to enhancing safety measures for teens on the Platform. This initiative responds to recent global research indicating that teens and their parents want more collaborative opportunities with online platforms. Created in partnership with Praesidio Safeguarding, the Youth Council consists of 15 teens from diverse backgrounds and countries, including the United States, United Kingdom, Brazil, Indonesia, Ireland, Kenya, Mexico, and Morocco. The group, which first convened in December 2023 and met again in February 2024 with TikTok Chief Executive Officer Redacted has prioritised teen well-being and inclusion for the year. They are contributing to the redesign of TikTok's Youth Portal and have sought clarity on what happens after an account or video is reported , aiming to ensure a safer and more inclusive environment on the app. The Youth Council's insights help TikTok refine its risk detection measures and enhance the overall user experience for young people.

TikTok is updating its Youth Guide (now renamed Youth Portal) to include new safety features, based on feedback from the Youth Council on reporting and blocking, and is introducing new GIFs to help Younger Users locate these features. Members of the Youth Council provided input and urged TikTok to share more information about reporting and blocking to better understand what happens after a report is made.

g. **Bullying**

TikTok engaged with 17 experts through Safety Advisory Council meetings and 1:1 engagements on their opinions on TikTok's bullying policies and how TikTok can better protect Younger Users from bullying harms. The consultations highlighted a number of key areas where minors are more or uniquely exposed to bullying compared to adults, particularly in the areas of revictimisation through resharing of bullying content and non-physical parental discipline.

h. **Multi-stakeholder working group on Younger Users' use of social media**

In 2023, TikTok joined the United Nations Violence Against Children's Office Protection through Online Participation working group (PoP). This global initiative aims to understand how children and youth use digital platforms to be safe. TikTok is an active member of PoP's working group (through participation in virtual meetings), and has also supported and attended in-person events that have taken place in Brussels, Paris, and Madrid.

9. RISKS RELATED TO AGE ASSURANCE

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014697

While Younger Users have the right to freedom of expression, access to information, and other fundamental rights linked to the use of the Platform, TikTok considers it proportionate to impose some restrictions on access for Younger Users. Expert consultations have supported TikTok's approach to identifying and mitigating risks related to Age Assurance.

a. Multi-stakeholder initiatives for Age Assurance best practices

TikTok has worked with Microsoft to establish a global multi-stakeholder initiative on age assurance. The aims of this initiative are to facilitate a holistic multi-stakeholder dialogue that explores the complexities and nuances around assessing age in a digital environment and to increase alignment on key principles relating to age assurance. To do this members of the group are drawn from other industry partners, leading safety and privacy organisations such as the children's organisation WeProtect Global Alliance and the privacy think tank, the Centre for Information Policy Leadership. The first milestone of this work was a convening (funded by TikTok) of relevant stakeholders at the end of March 2024 in London. The dialogue underscored the importance of balancing rights and risks, involving experts from various sectors to identify challenges and opportunities. Critical discussions centred on developing evidence-based, child-informed, and risk-based approaches to age assurance, ensuring robust protection mechanisms while enabling safe online participation for minors. The group is also establishing four Working Groups on key themes relating to age assurance, including law and regulation, horizon scanning, global and regional perspectives and risk assessments.

TikTok has engaged with national authorities and regulators, including the AEPD in Spain and the Ministry of Education in France, who are proposing various age assurance solutions. TikTok also consulted with Italy's Garante and AGCOM, and France's CNIL, and continues discussions with the NGO euCONSENT on their emerging solutions to age assurance technology. Additionally, TikTok is part of the Digital Trust & Safety Partnership, which has produced best practice documents for the sector.

b. TikTok's Safety Advisory Council and content classification

In 2023, during an in-person Safety Advisory Council summit with global Youth Safety and Well-being experts, the Child Safety team conducted a session titled 'Autonomy for Younger Users & Content Restrictions.' This session, focused on content classification, provided crucial insights that supported TikTok's existing content classification efforts by incorporating global, evidence-based perspectives. Key insights gathered included the importance of developing localised interventions and policies that account for different acceptability standards for various age groups (13-15 and 16-17) in different countries. Additionally, the session emphasised the need to support Younger Users in making informed decisions and learning while acknowledging that such users develop at different paces and stages.

10. CIVIC INTEGRITY: RISKS TO ELECTIONS AND CIVIC INTEGRITY

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TikTok recognises that 2024 has been and will continue to be a critical year for EU and global democratic processes. Citizens may use TikTok to explore and engage with political ideas and interact with their representatives. However, this activity gives rise to risks that election misinformation and/or misconduct is shared and amplified on the Platform, which could lead to real world effects harming democratic processes. TikTok conducts extensive and continuous engagements with external stakeholders, including fact-checkers, misinformation experts, law enforcement, and potentially impacted groups in order to identify and define Election Misinformation risks and to design and implement mitigations.

a. **Identifying emerging risks**

TikTok conducts regular consultations with its ESAC, comprising leading external experts on technology and content-related issues. Through its Fact-Checking Programme, TikTok implements various measures to identify emerging misinformation-related risks, including ad hoc policy consultations and trend analysis with fact-checking partners. Eleven fact-checking partners across the EU help TikTok to understand and identify localised risks to election integrity. Additionally, TikTok's Community Partner Channel (CPC) also supports risk identification, by onboarding NGOs who act as trusted flaggers to report suspected harmful content, including election misinformation. TikTok has created a rapid response system to streamline the exchange of information between civil society organisations, fact-checkers, and online platforms, as part of TikTok's commitments, as signatories of the CoPD. The rapid response system is a time-bound dedicated framework for cooperation among signatories during the 2024 European Parliament elections which allows non-platform signatories to flag time sensitive content, accounts or trends that may present threats to the integrity of the electoral process. TikTok has been a proactive participant in the Permanent Taskforce and subgroups including the ones related to Crisis Response, Elections and AIGC set up under the CoPD.

b. **Addressing election misconduct, hacked materials, and misrepresented sources**

TikTok updated its Community Guidelines in May 2024 following individual consultations with over 40 global experts in democracy, tech policy and freedom of speech. The update added new policies and enforcement regarding unverified elections claims, election misconduct, distribution of hacked materials, and misrepresented civic sources.

c. **Fact-checking at scale**

TikTok's Fact-Checking Programme is engaged to determine whether potential misinformation content that is not already contained in its database of fact- checked claims is in fact violative. This Programme involves 11 IFCN-verified fact-checkers in the EU, who assess unverified content based on their independent, expert, localised and linguistic knowledge. These fact-checkers include Agence France-Presse (AFP), dpa Deutsche Presse-Agentur (DPA), Demagog.pl, Facta, Faktograf, Logically Facts, Newtral, Maldita, Poligrafo, The Journal, Nieuwscheckers.

During 2023 and 2024, TikTok sought feedback and advice from its Fact-Checking Program members before updating its internal ratings labelling system in July 2024 for content that has undergone a fact checking process. This expanded the rating options available to fact-checkers so that they can label content as either 'false', 'misleading', 'unsubstantiated', 'inconclusive', 'true', or 'out of scope', which informs proportionate enforcement.

For detailed information on TikTok's Fact-checking Programme, and the media literacy campaigns launched in collaboration with these partners, please refer to Annex 5.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014699

11. CIVIC INTEGRITY: RISKS TO PUBLIC HEALTH FROM MEDICAL MISINFORMATION

TikTok held two consultations with academics and one with the World Health Organization (WHO) in December 2023 due to the potential global impact of Medical Misinformation. These consultations were valuable in helping frame TikTok's policy and enforcement updates for both critical and moderate harm health risks that launched in June 2024.

12. CIVIC INTEGRITY: RISKS TO PUBLIC SECURITY FROM HARMFUL MISINFORMATION CONTENT

TikTok recognises that misinformation can be highly localised in individual EU Member States with unique cultural, social, and political nuances. TikTok therefore leverages the insights and expertise of external stakeholders to develop effective and tailored strategies to combat misinformation that poses Public Security Risks on the Platform, ensuring that interventions are culturally sensitive and contextually relevant.

For detailed information on TikTok's Fact-checking Programme, and the media literacy campaigns launched in collaboration with these partners, please refer to Annex 5.

a. **Rapidly evolving events**

During events that may threaten public security, a vast amount of live or near live information can be shared rapidly. It is crucial to emphasise to users the importance of verifying information with official sources and the availability of tools to report violative content. In response to the unfolding Israel-Hamas war, TikTok collaborated with various fact-checking organisations including Agence France-Presse (AFP), Reuters and Fatabayyano as authoritative sources for conflict related information. The in-app search intervention that resulted from these partnerships was used by millions of users, and TikTok scaled the search intervention globally in March 2024 in response to the terrorist attack of the Crocus Hall in Russia.

b. **Climate change misinformation**

TikTok launched a $1m initiative to tackle climate misinformation across Spain, Brazil and the UAE during COP28. This Verified for Climate programme is a joint programme between the United Nations and Purpose, and brings together a team of verified champions, including scientists and trusted experts from Brazil, the United Arab Emirates, and Spain who supported select TikTok creators in developing educational content to tackle climate misinformation and disinformation. One such creator, Verified, developed a series of initiatives to assist with the dissemination of accurate climate change educational content as a result. TikTok then co-created campaigns to uplift the voices of creators within Verified's

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014700

network in order to increase public engagement at key moments, for example during UN Climate Week and COP28.

Verified for Climate partnered with 35 experts to produce 225 educational videos on TikTok, achieving 632+ million impressions and nearly 15.8+ million engagements globally. This partnership resulted in users gaining increased awareness and skills in identifying and reporting climate misinformation.

c. **Conferences to improve understanding of risk mitigation strategies**

TikTok attended multiple conferences and events to improve its risk mitigation strategies relating to misinformation, including the Global Fact-Checking Conference hosted by the International Fact-Checking Network in (Oslo, June 2022); The Trust & Safety Research Conference (Stanford University, September 2022); and the Association of Internet Researchers Conference (Technological University Dublin, November 2022).

13. FUNDAMENTAL RIGHTS: RISKS TO FUNDAMENTAL RIGHTS

TikTok is dedicated to engaging with stakeholders to uphold Fundamental Rights and address risks to such rights on and or due to its Platform. TikTok is committed to the UN Guiding Principles on Business and Human Rights and its call to conduct human rights due diligence across the lifecycle of its business. In this regard, TikTok has continued to engage in external partnerships with the objective of improving its ability to assess and mitigate risks to fundamental rights. TikTok has also maintained memberships in industry wide forums such as the Business for Social Responsibility's Human Rights Working Group.

a. **Balancing fundamental rights in risk assessment and mitigation**

TikTok regularly consults with Safety Advisory Council members in Europe, Asia, and Latin America to gain varied perspectives from civic society on best practices for the protection of fundamental rights. A key aspect of protecting these rights is navigating the delicate balance between competing rights and interests. TikTok conducted a dedicated consultation in June 2024 to seek advice and feedback on its risk assessment and management approach in light of the DSA's emphasis on fundamental rights and in particular, freedom of expression. TikTok consulted with Redacted a member of ESAC and Senior Legal Fellow at the Future of Free Speech. Redacted holds expertise in balancing content moderation with the protection of fundamental rights and the ongoing interest in TikTok's moderation and privacy measures. Redacted shared his view that the process of addressing and navigating the balance of fundamental rights is more important than the final position taken: for example, it can be acceptable to create a risk to freedom of speech due to mitigations against Hate Speech, so long as TikTok's process has meaningfully compared the two risks and assessed that the risk from Hate Speech is more severe than the risk to freedom of expression. This perspective underscores the importance of a deliberative process in policy-making regarding fundamental rights protection, which TikTok is committed to Redacted provided critical insights, highlighting how the DSA's definition of systemic risks may be interpreted at various levels - social, industry-wide, or specific to individual platforms. Joan recommended transparency about the complexities of risk assessments, advising against an over-

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014701

reliance and advocating for qualitative evaluations of risks to the Platform to ensure that a nuanced understanding is gained. Redacted also stressed the importance of engaging stakeholders from different contexts, languages, and cultures to tailor how various manifestations of fundamental rights risks are mitigated on the Platform. Redacted insights validated TikTok's methodological approach to systemic risk assessments under the DSA, highlighting TikTok's proactive approach to engaging with external stakeholders to identify and mitigate systemic risks.

b. Collaborating to anticipate future risks to fundamental rights

By participating in conferences and collaborating with peers and civil society stakeholders regarding fundamental rights, TikTok can share experiences and learn with peers to anticipate future risks.

In June 2023, TikTok experts attended the RightsCon global conference which convened business leaders, policy makers, general counsels, government representatives, technologists, academics, journalists, and human rights advocates from around the world to tackle pressing issues at the intersection of human rights and technology. TikTok also participated in RightsCon's 'Young leaders summit', in which students and young professionals engage in workshops to empower them to join policy and advocacy discussions around human rights in the digital age. The summit engaged these young leaders in advocacy, policymaking, and movement building for emerging issues, including data protection and privacy, network discrimination and connectivity, digital security, diversity and digital inclusion, and artificial intelligence and algorithmic accountability.

In June 2024, TikTok actively participated in the EU Rights & Risks Forum, an event hosted by Digital Trust and Safety Partnership and Global Network Initiative. The event brought together representatives from platforms, civil society, and academic experts from Europe and beyond. The forum focused on systemic risk assessments under the DSA and their implications for fundamental rights. Over two days of panels and workshops, TikTok, along with other participants, engaged in discussions aimed at informing and refining approaches to risk assessment and stakeholder engagement. The insights gained from this forum have contributed to TikTok's ongoing efforts in these areas.

TikTok's engagements with the WeProtect Global Alliance, Trust & Safety Professional Association, Tech Against Terrorism, Business for Social Responsibility, the Tech Coalition, and launch partners for the Partnership on AI's Framework Responsible Practices for Synthetic Media, mentioned in more detail elsewhere in this Annex, also all inform its approach to the mitigation of risks to fundamental rights.

14. AIGC

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014702

TikTok has collaborated with expert external stakeholders to develop TikTok's comprehensive Community Guidelines and policies covering the risks of AIGC.

c. **Fact-checking at scale**

TikTok's fact-checking partners are authorised to proactively identify to TikTok where they detected AIGC that may constitute Harmful Misinformation. TikTok is then able to take action to remove or label it (depending on the status/outcome of any necessary verification). Those partners will also provide TikTok with intelligence on prominent misinformation that is circulating on other social media platforms or websites that may benefit from verification.

d. **Election-related mis- and disinformation**

TikTok has joined industry partners as a party to the Tech Accord to Combat Deceptive Use of AI in 2024 Elections (the 'Accord'). Via the Accord, member technology companies are working together to safeguard communities against misleading and deceptive AI in this election year – including, with respect to AIGC, to safeguard against an increased risk of disinformation campaigns, efforts to discredit politicians and the proliferation of fake political endorsements by celebrities. Through this work, TikTok and industry partners aim to collaborate on tools to detect and address fake political online distribution of such AIGC, with attention to the importance of tracking the origin of deceptive election-related content and the need to raise public awareness about the problem.

e. **Transparency and responsible AI**

TikTok was the first social media platform signatory to the PAI's Responsible Practices for Synthetic Media (the 'Practices'), a first-of-its-kind code of industry best practices for AI transparency and responsible innovation, balancing creative expression with the risks of emerging AI technology. In accordance with its commitments as a launch partner for the practices, TikTok worked on a case study outlining how the Practices informed TikTok's policy making on synthetic media.

f. **Labelling AIGC**

TikTok is working to develop content provenance practices through implementation of the Coalition for Content Provenance and Authenticity (C2PA) standard in the second or third quarter of 2024. The C2PA standard is an open technical standard and content provenance solution that can provide information in the metadata for a piece of content about its origin and whether AIGC models were used to create or edit it. In 2024, TikTok collaborated with C2PA to integrate its standard into the Platform, enabling automatic AIGC labelling for off-platform content containing C2PA metadata. In 2023, TikTok held two consultations with media literacy experts and a behavioural scientist to develop language for an AIGC label, which was launched in September 2023. Additionally, TikTok conducted three external consultations on the use of satire in AIGC which informed updates to TikTok's synthetic and manipulated media policies, which were implemented in May 2024.

g. **Expert consultations**

To address and better understand the risks, trends, and opportunities that AIGC poses to youth, TikTok conducted consultations with experts like **Redacted**, and TikTok's ESAC members to guide TikTok's Youth Safety teams on the impact of existing (and new) trends on AIGC. TikTok participated in a roundtable discussion with AIGC experts put together by Ethical Resolve, and consulted with experts including **Redacted** (Data & Society and UC Berkeley); **Redacted** (Pennsylvania State University); **Redacted** (Bard College) **Redacted** (Yale Law School); **Redacted** (University of Virginia).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014703

ANNEX 5: OVERVIEW OF FACT-CHECKING PROGRAMME

Overview: Below is a high-level overview of TikTok's Fact-checking Programme (the "**Fact-checking Programme**"):

- **Context:** To deliver on its commitment to tackle Harmful Misinformation, including Election Misinformation, Medical Misinformation, and other Harmful Misinformation, TikTok's Integrity & Authenticity team has developed a Fact-checking Programme. This programme is led by members of the Trust & Safety team who have deep expertise and experience in Integrity & Authenticity issues, and entails a highly cross-functional approach with input from subject matter experts across various Trust & Safety teams (including our Integrity & Authenticity Policy, Product and Outreach Partnership Management ("**OPM**") teams), and multiple other teams;
- **Core objective and mission:** The core objective and mission of the Fact-checking Programme is to leverage the expertise of external fact-checking organisations to verify potential misinformation claims, and to enable our Trust & Safety teams to determine the appropriate enforcement action. This programme is grounded in our belief that technology companies should not alone make **decisions** about the veracity of borderline claims. For this reason, we consider that partnering with third-party fact-checking organisations is a critical component of our strategy to effectively tackle Harmful Misinformation;
- **Global framework for regional/local application:** The Fact-checking Programme provides a global overarching framework for fact-checking that can be applied in a consistent manner across a range of markets and languages. Since its launch, it has been expanded substantially to the point where we now work closely with 19 fact-checking partners globally, covering more than 100 countries and 50+ languages, including English, French, German, Spanish, Italian, Dutch, Austrian, Polish and Portuguese, along with a range of other languages. As part of the Fact-checking Programme, we can also secure temporary fact-checking resources when needed (and have done so in the past, for example, around elections).

1. OUR FACT CHECKING PARTNERS

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

a. **Our fact checking partners**

The fact-checking partners that we work with play a key role in contributing to the effectiveness and success of the Fact-checking Programme. TikTok therefore carefully considers the selection of which fact-checking partners we work with, based on those who have proven themselves to be reliable and capable. Within the EU, we partner with 10 fact-checking organisations who provide fact-checking coverage in 22 official EU languages: Agence France-Presse (AFP), dpa Deutsche Presse-Agentur, Demagog, Facta, Faktograf, Lead Stories, Logically Facts, Newtral, Poligrafo and Science Feedback. As part of the Fact-checking Programme, we can also secure temporary fact-checking resources when needed (and have done so in the past, for example, during EU Parliamentary elections expanding coverage for Maltese language). Our partners have specialists, including trained journalists, who review and verify reported content. Our moderators then use that independent feedback to take action and where appropriate, remove or make ineligible for recommendation false or misleading content or label unverified content. We have 3 primary criteria when selecting partners. More info here:

- **IFCN accredited:** The International Fact-Checking Network ("**IFCN**") is a global leader in fact-checking excellence that supports more than 100 fact-checking signatories around the world focused on best practices. Through the Fact-checking Programme, we only work with partners who are signatories to IFCN Code of Principles. Signatory status is based on an application process to the IFCN, and applicants are assessed by independent assessors for compliance with 31 criteria. Their assessment is reviewed by the IFCN Advisory Board to ensure fairness and consistency;
- **Relevant experience:** We generally choose partners who have experience working with technology companies, and especially those with deep expertise in fact-checking video content. This means that our fact-checking partners are a good fit for the types and formats of content that may require fact-checking as part of our Fact-checking Programme; and
- **Relevant language skills and support:** We prioritise partners with staff who have deep expertise in the markets they are covering, along with relevant language and topic knowledge.

b. **The fact-checking process**

As noted above, the core objective of the Fact-checking Programme is to leverage the expertise of external fact-checking organisations to determine enforcement actions on the most harmful and difficult to verify claims. The fact-checking process operates as follows:

- **Detection:** Where content is proactively detected or reported as being in violation of our Harmful Misinformation policies, it is routed and centralised to our dedicated specialised moderation team for integrity-related issues ("**Integrity Operations**") for review;
- **Integrity Operations Review:** Within our Trust & Safety Operations function, we have established a dedicated Integrity Operations team who moderate suspected misinformation content, which is made up of experienced moderators who then undergo specialised training on Integrity & Authenticity policy issues. Our Fact-Checking Programme is integrated into our specialised misinformation moderation workflow to inform our Integrity Operations team's decisions on whether content violates our misinformation policies;
- **Escalation:** If the Integrity Operations moderators are unsure as to the veracity of the video and believe it may contain misinformation, they have the ability to escalate to a fact-checking partner with expertise in the market who will assess the video, conduct reporting, and verify whether the information is false. Pending the fact-checker review, such videos will not be eligible for recommendation in the Platform's For You Feed. Our fact-checking partners access content which has been flagged for review through a dashboard made available for their exclusive use. The dashboard shows our fact-checkers certain quantitative information about the services they

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014705

provide, including the number of videos queued for assessment at any one time, as well as the time the review has taken. Fact-checkers can also use the dashboard to see the rating they applied to videos they have previously assessed;

- **Enforcement:** Our fact-checking partners do not take direct actions on the content that they review. After a fact-checker assesses a video, it returns the matter to the Integrity Operations team who then uses the fact-checker insights to determine whether the video violates our Harmful Misinformation policies. In terms of sanctions that may be imposed, violating videos can be removed globally, Not Recommended (which means they will not be eligible for recommendation in the Platform's For You Feed and we take measures to ensure that this content is harder to find in our search functionalities on the Platform), or labelled as unverified content. If content is confirmed to be accurate, and the content is in line with TikTok's Community Guidelines, it will remain on the platform;
- **Repository of fact-checked misinformation:** To ensure it is both efficient and scalable, as part of the Fact-checking Programme, our Integrity & Authenticity team maintains a repository of fact-checked information - such as claims, statements and conspiracy theories, and that have been fact-checked by our external fact-checking partners. This is a key resource for our Integrity Operations team to help them accurately, quickly and efficiently assess the veracity of suspected misinformation as part of their review process. This database is updated on an ongoing basis to include information that has been assessed and determined by our fact-checking partners to be verifiably false or misleading.

2. OTHER SUPPORT AND SERVICES

a. **Other support and services**

In addition to fact-checking services, we often work with our fact-checking partners where they are able to provide ancillary support to our Fact-checking Programme and misinformation mitigation strategy more widely, including:

- **Policy consulting/trends analysis:** This may involve our fact-checking partners providing us with reports identifying general misinformation trends observed on our Platform and across the industry generally, including new/changing industry or market trends, events or topics that generated particular misinformation or disinformation;
- **Media/digital literacy campaigns:** We can collaborate with fact-checking partners to produce media/digital literacy videos/campaigns to help inform and educate users about identifying and reporting misinformation. These videos can also educate users about specific topics such as civic engagement. For example, in partnership with our fact-checking partners, we ran integrity campaigns in advance of several European major elections in 2023 and 2024:
    - **Media Literacy: stand-alone campaigns in the EU:** Since August 2023, we rolled out an additional 13 new media literacy campaigns in Europe in collaboration with local media literacy bodies and our trusted fact-checking partners on topics identified as priority areas for increased digital literacy for our

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014706

users in Europe; for example, the Russia-Ukraine war or general critical thinking skills including identifying *misinformation* and *manipulated media and deep fake*. Of these media literacy campaigns:

- Six (**6**) campaigns launched in Austria, Bulgaria, Czech Republic, Croatia, Germany and Slovenia were specific to the war in Ukraine, including identifying misinformation and manipulated media reports;
- Seven (**7**) campaigns launched in Finland, Ireland, Italy, Spain, Sweden, Denmark and Netherlands were focused on general media literacy and critical thinking skills to fight misinformation and manipulated media online, H5 Example below.

b. **Other support and services**

In addition to stand-alone media literacy campaigns, starting from 2023 and throughout 2024, we introduced media literacy resources in our Elections Centres including tips about how to develop critical thinking skills, identifying *misinformation, manipulated media and deep fake*. The media literacy section in our elections centres includes videos from our trusted fact-checking partners (see table below):

- **2023 Slovak Parliamentary Election:** From 4 September 2023, we launched an in-app Election Centre in collaboration with electoral commissions to provide users up-to-date information about the 2023 Slovak parliamentary elections. The centre contained a section about fighting-fake news, including videos;
- **2023 Polish Parliamentary Election:** From 18 September 2023, we launched an in-app Election Centre in collaboration with electoral commissions to provide users up-to-date information about the 2023 Polish parliamentary elections. The centre contained a section about fighting-fake news, including videos created in partnership with media local media literacy organisation and fact-checkers Demagog.pl and FakeNews.pl;
- **2023 Dutch General Elections:** From 25 October 2023, we launched an in-app Election Centre to provide users up-to-date information about the 2023 Dutch House of Representatives elections. The centre contained a section about fighting-fake news, which linked to media literacy association isdatechtzo.nl and included videos created in partnership with fact-checking organisations dpa Deutsche Presse-Agentur and Nieuwscheckers;
- **2024 Finnish Presidential Election:** From 1 January 2024, we launched an in-app Election Centre to provide users up-to-date information about the 2024 Finnish presidential elections. The centre contained a section about fighting-fake news, which linked to media literacy association Mediataitokoulu and included videos created in partnership with fact-checking organisation Logically Facts;
- **2024 Irish Referendum:** From 2 February 2024, we launched an in-app Referendum Centre to provide users up-to-date information about the 2024 Irish Referendum and legislative elections. The centre contained a section about fighting-fake news, which linked to media literacy association medialiteracyireland.ie and included videos created in partnership with fact-checking organisations The Journal and Logically Facts;
- **2024 Slovak Presidential Election:** From 3 March 2024, we launched an in-app Election Centre to provide users up-to-date information about the 2024 Slovak presidential elections. The centre contained a section about fighting-fake news, including videos created in partnership with media local media literacy organization DigiQ;

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014707

- **2024 EU Parliamentary Election:** From 28 March 2024, we launched an individual in-app Election Centre for each EU member state available in 27 EU Languages. Working with electoral commissions, European Parliament, civic society organisations including Fact-checkers, these Centres connected people with reliable voting information, including when, where, how to vote and, ultimately, the election results themselves. Between March and June 2024, a few countries in the EU, including Croatia, Lithuania, Bulgaria and Belgium, were called to national wide elections. In this specific instances we worked with national electoral commissions to include relevant voting information in our in-app EU Election Centre. All 27 Elections Centres contained a media literacy section that includes videos from our trusted fact-checking and partners local media literacy bodies: Agence France-Presse (AFP), dpa Deutsche Presse-Agentur (DPA), Demagog.pl, Demagog.cz, Facta, Faktograf, Logically Facts, Newtral, Poligrafo, Delfi.lt, The Journal, Nieuwscheckers, Funky Citizens, DigiQ, Ostro; and
- **2024 French Parliamentary Election:** From 17 June 2024, we launched an in-app Election Centre to provide users up-to-date information about the 2024 French Legislative elections. The centre contained a section about fighting-fake news, including videos created in partnership with fact-checking organisations Agence France-Presse (AFP).

c. Knowledge development

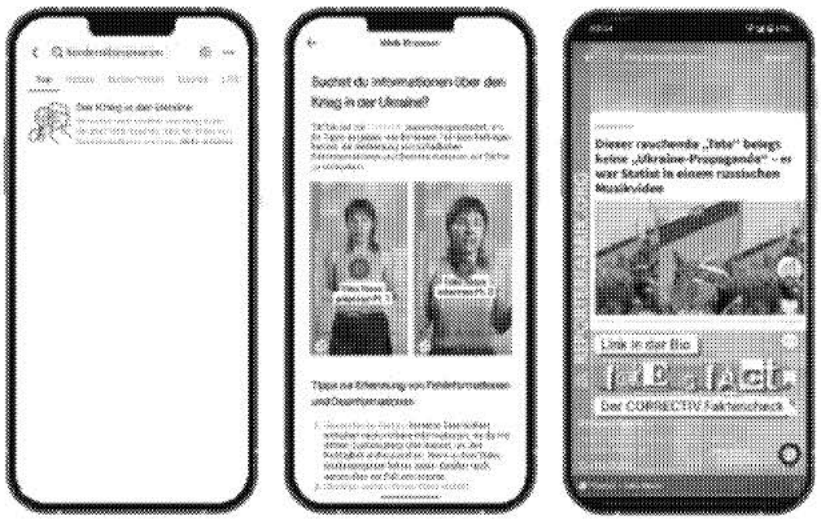
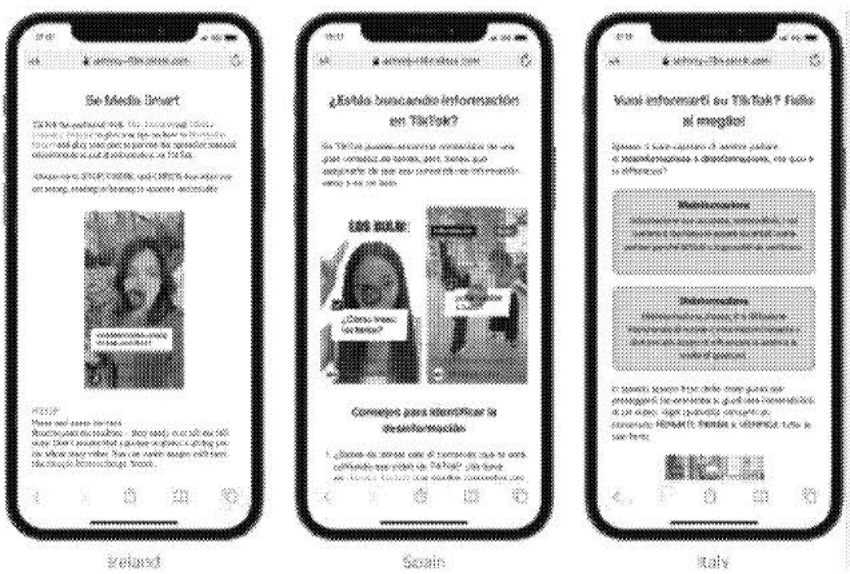
In order to effectively scale the expertise and feedback provided by our fact-checkers globally, we have implemented the measures listed below:

- **Fact-checking repository:** As noted above, we have developed a database of claims which have previously been fact-checked by our global fact checking partners. This contributes to knowledge development and achieves efficiencies because feedback from fact-checkers is more easily scalable and, therefore, avoids our moderators repeatedly sending the same claims for review and verification, and enables them to make swifter decisions on suspected misinformation content;
- **Proactive detection by our fact-checking partners:** Each of our fact-checking partners are authorised to proactively identify content that may constitute Harmful Misinformation on our Platform and suggest prominent misinformation that is circulating on other social media platforms or websites that may benefit from verification; and
- **Fact-checking guidelines:** We create guidelines and trending topic reminders for our moderators on the basis of previous fact-checking assessments. This ensures our moderation teams leverage the insights from our fact-checking partners and helps our moderators make swift and accurate decisions on flagged content regardless of the language in which the original claim was made.

3. TIKTOK MEDIA LITERACY CAMPAIGNS IN EUROPE	
Election Campaign	**Relevant TikTok Interface**

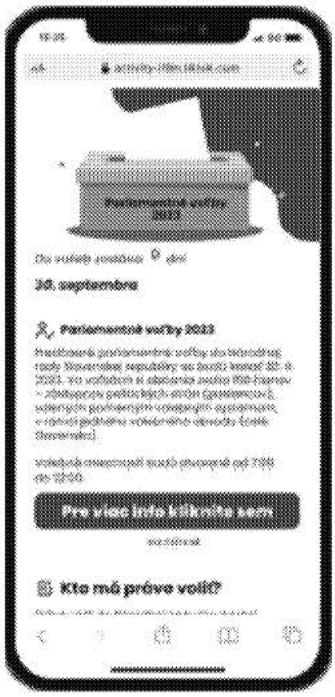
CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014708

Six (**6**) campaigns launched in Austria, Bulgaria, Czech Republic, Croatia, Germany and Slovenia were specific to the war in Ukraine, including identifying misinformation and manipulated media reports.	 Germany and Austria
Seven (**7**) campaigns launched in Finland, Ireland, Italy, Spain, Sweden, Denmark and Netherlands were focused on general media literacy and critical thinking skills to fight misinformation and manipulated media online, H5 Example below.	 Ireland Spain Italy

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014709

2023 Slovak Parliamentary Election: From 4 September 2023, we launched an in-app Election Centre in collaboration with electoral commissions to provide users up-to-date information about the 2023 Slovak parliamentary elections. The centre contained a section about fighting-fake news, including videos	
2023 Dutch General Elections: From 25 October 2023, we launched an in-app Election Centre to provide users up-to-date information about the 2023 Dutch House of Representatives elections. The centre contained a section about fighting-fake news, which linked to media literacy association isdatechtzo.nl and included videos created in partnership with fact-checking organisations dpa Deutsche Presse-Agentur and Nieuwscheckers.	

TikTok

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

2024 Finnish Presidential Election: From 1 January 2024, we launched an in-app Election Centre to provide users up-to-date information about the 2024 Finnish presidential elections. The centre contained a section about fighting-fake news, which linked to media literacy association Mediataitokoulu and included videos created in partnership with fact-checking organisation Logically Facts.	
2023 Irish Referendum: From 2 February 2024, we launched an in-app Referendum Centre to provide users up-to-date information about the 2024 Irish Referendum and legislative elections. The centre contained a section about fighting-fake news, which linked to media literacy association medialiteracyireland.ie and included videos created in partnership with fact-checking organisations The Journal and Logically Facts.	

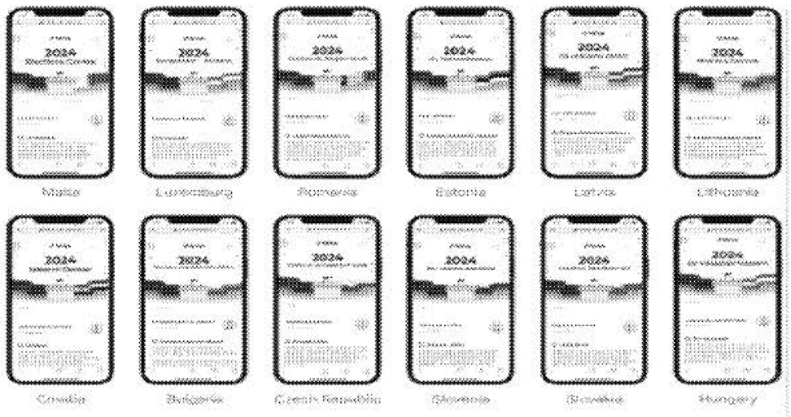
CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014711

2023 Slovak Presidential Election: From 3 March 2024, we launched an in-app Election Centre to provide users up-to-date information about the 2024 Slovak presidential elections. The centre contained a section about fighting-fake news, including videos created in partnership with media local media literacy organisation DigiQ.	
2024 EU Parliamentary Election: From 28 March 2024, we launched an individual in-app Election Centre for each EU member state available in 27 EU Languages. Working with electoral commissions, European Parliament, civic society organisations including Fact-checkers, these Centres connected people with reliable voting information, including when, where, how to vote and, ultimately, the election results themselves. Between March and June 2024, a few countries in the EU, including Croatia, Lithuania, Bulgaria and Belgium, were called to national wide elections. In these specific instances we worked with national electoral commissions to include relevant voting information in our in-app EU Election Centre. All 27 Elections Centres contained a media literacy section that includes videos from our trusted fact-checking and partners local media literacy bodies: Agence France-Presse (AFP), dpa Deutsche Presse-Agentur (DPA), Demagog.pl, Demagog.cz, Facta, Faktograf, Logically Facts, Newtral, Poligrafo, Delfi.lt, The	

TikTok

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Journal, Nieuwscheckers, Funky Citizens, DigiQ, Ostro.	

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Exhibit 295

TikTok Technology Limited's response to RFI
(Ref.Ares(2024)6974764 - 02/10/2024)
2 October 2024

Introduction

TikTok Technology Limited submits this response (the "**Response**") to the Commission's Request for Information Ref. Ares(2024)6974764 (the "**RFI**") as the legal entity that is the provider of the designated very large online platform ("**VLOP**") TikTok (the "**Platform**") in the EU under Regulation (EU) 2022/2065 (the "**DSA**").

Questions and Answers

1. In TikTok's 2024 risk assessment report, "*the impact of recommender systems on the content that users consume*" has been identified as a "*priority factor*" posing "*potential risks across many of its assessments*" ([1]). However, the report provides only limited information on how the provider of TikTok analysed and assessed whether and how the design or the functioning of its different recommender systems influences such risks. In this context, please provide the following information:

a. Pursuant to Article 34(3) of Regulation (EU) 2022/2065, please provide all the risk assessment "Modules" underlying the 2024 risk assessment report.

1. TikTok understands that, by its request for TikTok to provide the Year 2 "Modules", the Commission is in practice seeking to obtain the detailed analysis undertaken by TikTok in the Art. 34 DSA risk assessment process underlying its (updated) Year 2 Risk Assessment Report dated 2 October 2024 (the "**Year 2 RAR**") (enclosed as Annex 001).

2. As explained in TikTok's Year 2 RAR[2] and TikTok's Art. 34 DSA Risk Assessment Guidelines dated 25 August 2023 (enclosed as Annex II to TikTok's response to Section 1 of RFI Ref.Ares(2024)1961604 (the "**GenAI RFI**")), and as also explained in TikTok's updated Art. 34 DSA Risk Assessment Guidelines dated 28 August 2024, TikTok used the term "Module" to refer to distinct areas of systemic risks and the risk assessment exercises that TikTok conducted for each of those specific areas.

3. Each RAR summarises the results of TikTok's detailed Art. 34 DSA risk assessments (as explained in the Year 1 Risk Assessment Report dated 29 September 2023 (the "**Year 1 RAR**") and Year 2 RARs in their respective "*How to use this report*" annexes). The RARs are not intended as a comprehensive or exhaustive overview of the detailed analysis and underlying assessment. For Year 1, the detailed analysis and assessments were recorded in specially prepared forms, corresponding to each Module ("**Module Reports**"). Those Module Reports were previously provided to the Commission.

4. With its Year 1 risk assessment, TikTok established a "baseline" (as also explained in the Year 2 RAR[3]). In Year 2, TikTok built on that baseline, consistent with Recital 85 DSA (which provides that subsequent risk assessments should build on each other and show the evolution of the risks identified). To help to facilitate that cumulative approach and to efficiently and consistently identify the evolution of risks, TikTok adapted its risk assessment methodology for Year 2.

5. In Year 2, TikTok has completed:

    a. a detailed questionnaire for each Module (i.e. area of systematic risk), which were completed by the relevant subject matter experts, and which were designed to facilitate and prompt diligent identification and assessment of both new and evolving risks by those

[1] See: "*TikTok DSA risk assessment report 2024*" submitted by the provider of TikTok to the European Commission on 29 August 2024, registered under Ares(2024)6149012, pp. 2 and 7.
[2] pp. 3-6.
[3] E.g., on pp. 2 and 5-6.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

subject matter experts, and to record a summary of their responses, including data relied on, ("**Questionnaires**");

b. a risk scoring rubric, which was used for the risk scoring calculations referred to at pages 5-8 of the Year 2 RAR (the "**Risk Scoring Rubric**") and incorporates elements of the ISO31000 risk scoring methodology; and

c. minutes of meetings held by each Module's specialist Risk Assessment Review Group made up of senior stakeholders ("**RARGs**"), recording the outcome of the risk assessment-related discussions and related decisions taken during those meetings, in particular on the basis of the Questionnaires and Risk Scoring Rubric ("**RARG Minutes**").

6. The areas of systemic risks (Modules) remained substantially the same between TikTok's Year 1 RAR and its Year 2 RAR. The Modules / risk areas are:

a. CSAM;
b. Age Assurance;
c. Age-Appropriate Content;
d. Online Engagement;
e. Elections & Civic Integrity;
f. Gender-based Violence;
g. Terrorist Content;
h. Hate Speech;
i. Medical Misinformation;
j. Harmful Misinformation;
k. Fundamental Rights; and
l. IP-Infringing Content.

7. However, TikTok understands that the Commission asks TikTok to produce the documents that contain the detailed analysis and underlying assessment. TikTok therefore encloses the following documents:

a. the Risk Scoring Rubric (as Annex 002); and

b. corresponding to each of the Module Reports:

i. all the Questionnaires (as Annexes 003 - 014); and

ii. all the RARG Minutes (as Annexes 015 - 026),

underlying TikTok's Year 2 RAR.

b. Pursuant to Article 34(3) of Regulation (EU) 2022/2065, please provide all supporting documents identifying, analysing and assessing whether and how all of TikTok's recommender systems and other relevant algorithmic systems influence risks to electoral processes, civic discourse and public security, illegal hate speech, gender-based violence, and fundamental rights, including the pluralism of the media.

8. TikTok understands that:

a. the Commission's reference to "*electoral processes, civic discourse and public security, illegal hate speech, gender-based violence, and fundamental rights, including the pluralism of the media*" is referring to the following Modules: (i) Elections & Civic Integrity; (ii) Harmful Misinformation; (iii) Hate Speech; (iv) Gender-based Violence; and (v) Fundamental Rights (the "**Relevant Modules**"); and

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

b. by its request, the Commission is seeking to obtain the supporting documents: (i) relating to TikTok's Year 2 risk assessment under the Relevant Modules; and which (ii) identify, analyse and assess whether and how all of TikTok's recommender systems, and other algorithmic systems relevant to its recommender systems, influence the risks assessed under the Relevant Modules.

9. Copies of the supporting documents that TikTok has identified as responsive to the Commission's request are provided as Annexes 027 - 040.

c. Please provide a copy of the presentation of the 2024 risk assessment report to the provider of TikTok's board of directors.

10. A copy of the presentation of the Year 2 RAR to TikTok's board of directors is provided as Annex 041.

11. Following the board meeting on 22 August 2024, the RARG members reviewed the risk scoring of the election integrity and general misinformation risks and the underlying data, as suggested at the board meeting, and determined that the likelihood of the risk should be reassessed.

12. Following that reassessment, the likelihood risk score was increased from Tier 3 to Tier 4. This changed the residual risk from "*low*" to "*moderate*". The higher risk score shows that this was identified as an area of priority.

d. Please provide a non-technical description of the differences between the recommender systems currently deployed by the provider of TikTok and publicly available recommendation algorithms, architectures, and techniques, such as Monolith ([4]), CowClip ([5]), and HLLM ([6]).

13. Monolith, CowClip, and HLLM are not deployed in the recommender systems on the TikTok Platform. Accordingly, TikTok is not in a position – nor does it consider that it has an obligation under Article 34 DSA – to analyse any differences between those algorithms, architectures, and techniques and the recommender systems deployed by TikTok on its Platform.

14. For the Commission's reference:

a. Monolith is a business-to-business solution to provide an intelligent recommendation platform to enterprise customers. TikTok's use case is business-to-customer, whereas Monolith is business-to-business.

b. CowClip is a general technique for predicting whether a user will click on a recommended item. It was published in an academic journal (https://arxiv.org/pdf/2204.06240).

c. HLLM was also an academic project (published here: https://arxiv.org/pdf/2409.12740), with the goal of applying large language models ("**LLMs**") to recommender systems.

---

[4] See: Liu et al. (2022). Monolith: Real Time Recommendation System With Collisionless Embedding Table. In Proceedings of 5th Workshop on Online Recommender Systems and User Modeling, in conjunction with the 16th ACM Conference on Recommender Systems (ORSUM@ACM RecSys 2022). ACM, New York, NY, USA.

[5] See: Zheng et al. (2023). CowClip: Reducing CTR Prediction Model Training Time from 12 Hours to 10 Minutes on 1 GPU. Proceedings of the AAAI Conference on Artificial Intelligence, 37(9), 11390-11398. https://doi.org/10.1609/aaai.v37i9.26347, accessed on 26 September 2024.

[6] See: Chen et al. (2024). HLLM: Enhancing Sequential Recommendations via Hierarchical Large Language Models for Item and User Modeling. arXiv preprint. https://arxiv.org/abs/2409.12740 accessed on 26 September 2024.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013691

2. In TikTok's 2024's risk assessment report, the provider of TikTok identifies that "*there is an inherent risk that features and advancements in content moderation models may be developed in a biased or discriminatory manner. This risk arises from the potential development of these features and policies using insufficiently inclusive data sets, safeguard and safety controls*" ([7]). Please explain whether the provider of TikTok analysed and assessed whether potentially biased content moderation systems may impact recommendations and the functioning of the specific recommender systems.

15. TikTok understands that, in this ***question 2***, the Commission is asking specifically whether, as part of its Year 2 risk assessment, TikTok analysed and assessed whether the inherent risk that content moderation systems may be developed in a biased or discriminatory manner due to insufficiently inclusive data sets, safeguards, and safety controls (the "**Inherent Risk**")[8] may impact recommendations and the functioning of recommender systems on the Platform (the "**Relevant Risk**").

16. Content moderation systems are separate and distinct systems from recommender systems, and therefore they (and the Inherent Risk) do not (and cannot) impact on the *functioning* of recommender systems themselves (see further TikTok's response to ***question 1(d)*** above).

17. However, TikTok's content moderation systems assess, and may classify and/or carry out content moderation on, the pool of content that may potentially be recommended by TikTok's recommender systems. Therefore, the content moderation systems may impact the *recommendations* of content to users.[9]

18. In response to the Commission's question, TikTok did analyse and assess the Relevant Risk as part of its Year 2 risk assessment process, as explained further below.

19. TikTok is aware of the inherent relationship between content moderation systems and recommender systems, and the potential impact of the former on the recommendations (as explained in paragraph 17 above). TikTok's risk assessments also took that relationship into account:

    a. For example, in TikTok's Year 1 risk assessment process (which, as explained in response to ***question 1(a)*** above, TikTok's Year 2 risk assessment process built on), TikTok's Elections & Civic Integrity Module Template Report explains that mitigation measures "*to prevent Election Misinformation from being recommended or surfaced in a user's FYF*" include:

        i. "*where* [Election Misinformation] *content is detected or reported as being in violation of our harmful misinformation policies, it* ***gets reviewed*** **[and may be]** ***removed****... or* ***deemed not eligible for recommendation in the FYF***"; and

        ii. (in relation to "***Content removal (or ineligible for recommendation)***"): "*Where content is found to violate TikTok's Community Guidelines on Civic and Election Integrity* ***through*** **[TikTok's]** ***content moderation processes****... it is removed from the Platform, except for content that is labelled as* ***ineligible for recommendation*** *in the FYF... Due to the emphasis that TikTok places on proactively detecting and moderating harmful content, this* ***mitigates the risk of Election Misinformation being recommended or surfaced in a user's FYF***"; and

    b. In its Year 2 Fundamental Rights Module Questionnaire, TikTok explained that: "*We have also provided notice to users on when and* ***why their content is demoted or removed from***

---

[7] See: "*TikTok DSA risk assessment report 2024*" submitted by the provider of TikTok to the European Commission on 29 August 2024, registered under Ares(2024)6149012, p. 80.

[8] For an explanation of how TikTok assesses inherent risks and residual risks, see pp. 7-8 of the Year 2 RAR

[9] As impliedly recognised in the definition of 'content moderation' under Article 3(t) DSA, which includes, among other things: (i) "*measures taken that affect the... visibility* [of illegal content]", which per Recital 55 DSA "*may consist* [of] *demotion in ranking or in recommender systems*"; and (ii) "*demotion*" of such content, i.e. in recommender systems' rankings.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013692

***recommendations**, following the Santa Clara Principles* [On Transparency and Accountability in **Content Moderation**]... ".

20. More generally, TikTok's Trust and Safety Platform Fairness ("**Platform Fairness**") team leads a program to address potential bias and discrimination in the context of content moderation. The team provides input on policy design to consider principles of fairness, inclusion, and human rights. This includes undertaking assessments and consultations to ensure that policy designs are fair and inclusive.

3. Search suggestions (e.g., 'Others searched for') on TikTok are becoming a relevant source of information, especially for young users, and particularly in the context of electoral processes. In this respect, recent research indicates how search suggestion systems employed on TikTok may amplify, among others, risks for electoral processes and civic discourse ([10]). Please provide the following information:

a. How has TikTok identified, analysed and assessed the extent to which risks for elections and civic discourse are impacted by the design or functioning of its search suggestion systems?

21. TikTok understands that, in this ***question 3(a)***, the Commission is asking about the following search suggestion features on the Platform (together, "**Search Suggestions**"):

    a. "**Suggested Search Terms**" – this feature suggests search terms (or 'queries') which expand upon a term the user has started to enter. For example, in response to the user entering "*cha*" into the Search bar, Suggested Search Terms might suggest longer search terms such as "*cha cha dance*", "*charcuterie board*", and "*charms for bracelets*";

    b. "**Feedbar**" – this feature suggests searches after the user watches a video and appears in a search bar at the bottom of a video;

    c. "**Comment Words**" – this feature suggests searches within the comment panel of a video, and can appear in the ceiling of the comment panel or as highlighted words within individual comments;

    d. "**Video Page Search Bar**" – this feature suggests additional related searches while the user scrolls through videos from the search results page. Video Page Search Bar appears at the top of the videos;

    e. "**Others Searched For**" – this feature recommends additional searches based on content that the user just searched for. The feature may appear in the search results page itself or while the user scrolls through videos from the search results page; and

    f. "**You May Like**" – this feature recommends searches to the user within the search bar before they enter a search term, which are not limited by reference to recent searches run or content viewed.

22. The pool of potential search suggestions that TikTok may recommend to users is primarily based on search terms used by users on the Platform. In some Search Suggestions features, such as "You May Like", TikTok may also base search suggestions on the accounts of people the user may know. Prior to TikTok recommending search suggestions to users, they are subject to moderation measures, as discussed in TikTok's response to ***question 3(b)***.

[10] For example, as reported by Interface and AI Forensic, TikTok's search suggestion feature "*suggests oddly specific and partially misleading search ideas that can lead users to believe questionable information or send them down contentious political rabbit holes*". See: https://tiktok-audit.com/blog/2024/Search-Suggestions/, accessed on 26 September 2024; See also: https://crossover.social/wp-content/uploads/2024/07/TheTiktokEffect.pdf, accessed on 26 September 2024.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

23. With respect to the trend referred to in the Commission's introduction to ***question 3***, TikTok is aware that users have used TikTok's Search Suggestions in recent years to search for content, including political and other elections-related information. TikTok considers this to be associated with changes in user behaviour.

24. TikTok considered in its Article 34 risk assessment process that – because recommender systems (which include Search Suggestions) suggest or prioritise specific information or content to users of online platforms – they may potentially play a role in the amplification and dissemination of illegal or harmful content related to civic discourse and electoral processes (see, e.g., the Elections & Civic Integrity Module section of the Year 2 RAR). Accordingly, absent appropriate risk mitigation measures (which TikTok has put in place, as described in response to ***question 3(b)*** below), there is an inherent risk that recommender systems (including Search Suggestions) may recommend such content.

25. TikTok also considers this inherent risk in the context of its Elections Integrity Programme, which is designed to ensure that TikTok can anticipate, prepare for, and prevent or mitigate relevant risks in relation to elections integrity. As part of the Elections Integrity Programme, TikTok undertakes scenario planning to anticipate the novel and time-sensitive policy issues which elections may present. This involves TikTok assessing how the identified risks could manifest themselves on the Platform in the context of an election, across a range of areas – including at a policy level, at a product level, and in respect of TikTok's processes.

26. TikTok has created, as part of its Elections Integrity Programme, an internal standard operating procedure policy ("**SOP**") to help it proactively identify and address potentially problematic Search terms that may be recommended to users, including through Search Suggestions such as Suggested Search Terms and Others Searched For. The SOP provides: (i) a proactive strategy toolkit for TikTok, with tailored solutions to address emerging Search-related risks; and (ii) a flexible and readily deployable framework to streamline TikTok's risk management processes for elections to help promote efficient responses to relevant risks.

27. In the SOP, TikTok assessed and identified relevant risks as including:

    a. Heightened risks from high-profile elections: High-profile elections pose substantial risks due to heightened emotions, intense public engagement, and the potential for election-related misinformation and polarising content to spread rapidly. The volatile nature of political discussions during these events can lead to the amplification of harmful narratives and other negative outcomes.

    b. Distinct risks of Search Suggestion features during elections: Suggested Search Terms and other Search Suggestions during elections introduce distinct risks that, if left unmitigated, could inadvertently amplify verifiably false or misleading content related to an election or other civic events in the EU (such as referenda or a census) ("**Election Misinformation**"), biassed viewpoints, or contentious content, potentially affecting public sentiment and electoral results. TikTok identified that a proactive strategy is necessary to mitigate Election Misinformation, particularly in Suggested Search Terms.

    c. Challenges with keywords as a mitigation method: Given the intricate landscape of risks associated with elections – in particular high-stakes elections – developing keyword blocklists (see further TikTok's response to ***question 3(b)*** below) may not be sufficient to effectively address the relevant risks.

28. Further, TikTok conducted specific risk assessments related to the elections for the European Parliament which took place between 6-9 June 2024 (the "**EU Elections**"):

    a. Scenario assessment and response planning: With a particular (but not exclusive) focus on Suggested Search Terms and Others Searched For, TikTok identified and assessed certain specific risk scenarios ahead of the significant elections which took place globally in 2024, including the EU Elections. TikTok assessed each scenario's likelihood and risk level. For each identified risk scenario, TikTok developed a specific response plan (including steps

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

such as additional content moderation reviews, fact-checking and involvement of TikTok's specialist internal teams).

b. "Priority markets" assessment: TikTok also conducted specific assessments of risks in Spain, France, Italy, and Germany to assess the risk levels associated with searches of topics and candidates related to the EU Elections.

c. EU Trust and Safety Risk Assessment: Additionally, in January 2024, TikTok's Trust and Safety Team conducted a further assessment of potential risks ahead of the EU Elections. In this process, TikTok specifically identified and considered that young people are increasingly using TikTok Search as one of their primary news sources. TikTok's risk assessment also took into account the increased emphasis on young people voting in the EU Elections and that the voting age had been lowered in certain EU Member States.

29. On an ongoing basis, TikTok also produces internal reports to document and track Search-related risk issues that TikTok identifies and addresses, including in relation to Search Suggestions and Election Misinformation. The risks included in this reporting encompass those identified through external escalations, internal reports from TikTok's cross-functional teams, user reporting and feedback, and other sources. Through this reporting, TikTok seeks to record lessons learned, identify key areas for improvement, and outline ongoing and future action items. TikTok uses these insights to continuously enhance its risk assessment and risk management processes in relation to Search (including Search Suggestions) and build on its experience in this highly nuanced and complex area.

30. TikTok has implemented mitigation measures specifically addressing the risks related to Search Suggestions – as explained in response to ***question 3(b)*** below – which also reflect the risks identified, analysed and assessed by TikTok stemming from the design or functioning of those systems.

b. Which mitigation measures has TikTok put in place to mitigate risks for election and civic discourse that may arise from the design or functioning of its search suggestion systems?

31. TikTok understands that, in this ***question 3(b)***, the Commission is inquiring specifically about Search Suggestions, as defined in response to ***question 3(a)*** above. In relation to Search Suggestions, and in addition to its broader Search-related risk mitigation measures, TikTok has implemented a variety of measures to mitigate risks for election and civic discourse:

32. Content moderation of Search Suggestions and enforcement of Community Guidelines:

a. As explained in the Elections & Civic Integrity Module Report,[11] TikTok designed Search, including Search Suggestions, with a 'safety-first' approach. As part of this, automated safety models identify and filter out (potentially) harmful Search Suggestions through machine learning algorithms. TikTok uses different models to assign a risk score: (i) a natural language processing model ("**NLP**") and a large language model ("**LLM**"), which assess risk based on language; and (ii) a "multimodal" model, which can also take into account other information such as images. The models assess whether terms are likely to be violative of TikTok's Integrity & Authenticity (and other) policies, including relating to Misinformation and Civic and Election Integrity. Where Search Suggestions (and/or the content they are likely to produce) are determined to be violative, those will not be recommended. Prior to the EU Elections, TikTok launched a version of an LLM for Search Suggestions, with the ability to make even more nuanced assessments, in certain markets assessed to be high risk/high priority.

b. As a further mitigation measure in respect of Search Suggestions, TikTok enqueues the top-recommended Search Suggestions from the past two hours for a further automated

[11] p. 31.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013695

and/or human review. If those searches are rejected following that further review, they will no longer be recommended to users as Search Suggestions.

33. Keyword lists and blocked search results:

a. As referred to in TikTok's Elections & Civic Integrity Module Report,[12] TikTok uses keywords lists (which may include emojis and special characters) to supplement the capabilities to detect videos or other content that contain text, embedded text, audio transcripts or captions linked to Election Misinformation. If a user enters a search term that matches violative keywords on TikTok's list associated with Election Misinformation, then TikTok does not return Search Suggestions (or Search results). Instead, TikTok surfaces an information and resource page to educate the user on why the search terms are banned and may re-direct such users to authoritative election information and resources. Ahead of and during an election period, TikTok may supplement the list with keywords related to voter suppression, voter fraud, and election disinformation. For example, prior to the EU Elections, TikTok updated its keywords list, which was reviewed by TikTok's regional teams across all EU markets.

b. TikTok has processes in place to review queries and update its existing sensitive keyword list and its models to effectively assess risk that a specific search term may lead to content that contains Election Misinformation or otherwise violates the Community Guidelines. TikTok's intervention tools also help to identify keywords that are similar to keywords that have been detected and identified as violative. Political phrases, "dog whistle" terms, and imagery are constantly evolving. TikTok's automated moderation system (and human moderators) are frequently updated and trained with these market-specific terms. TikTok's cross-functional teams work as part of the Elections Integrity Programme to ensure its keyword lists are kept up to date.

c. TikTok also has protocols in place to quickly detect and control viral trends featuring harmful content, such as Election Misinformation (see further TikTok's response to ***questions 8*** and ***9*** below). Such risks can arise where seemingly harmless hashtags or search terms are misappropriated by bad actors and become associated with harmful user content. Potential mitigation measures include banning certain hashtags and/or blocking searches associated with emerging risks.

34. Election labels / Search banners: As part of its Elections Integrity Programme, and as described in TikTok's Year 2 RAR[13] and Elections & Civic Integrity Module Report,[14] TikTok also applies various "user education measures" that mitigate risks from search, including Search Suggestions. Those include Search banners. A Search banner is a graphic banner that appears as the top level search result for a given set of keywords, which is clickable and leads to an in-app destination (e.g., to the relevant Election Centre – see further TikTok's response to ***question 4*** below). TikTok also applies a 'Know the Facts' user-facing banner to content if a Fact-checking partner is unable to conclude whether a piece of content is Election Misinformation or not.

35. Account verification badge: TikTok displays the "verified badge" alongside verified accounts (which may include GPPPAs[15]) that are recommended in Search Suggestions, to help protect users from impostor accounts that may attempt to spread Election Misinformation while, e.g., pretending to represent a political party or a politician. See further TikTok's responses to ***question 5*** below.

The Report and Blogpost cited by the Commission

---

[12] p. 21.
[13] p. 67.
[14] pp. 49 and 68.
[15] i.e., Government, Politician, and Political Party Accounts, as explained at https://support.tiktok.com/en/using-tiktok/growing-your-audience/government-politician-and-political-party-accounts and addressed further in response to ***question 5*** below.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

36. TikTok does not propose to respond comprehensively to the "*The TikTok Effect*" report (the "**Report**") and the "*Other search for ... the opposition party*" blogpost (the "**Blogpost**") in this Response, but would be happy to discuss them further with the Commission separately.

37. However, TikTok wishes to note that:

    a. The Report and Blogpost do not appear to support the broad and generalised statements in the introduction to this ***question 3***.

    b. TikTok's comments in ***question 5*** below, regarding Articles 8 and 14 DSA, are similarly applicable in the context of this ***question 3***. Further, it is not possible for content moderation (and the removal of Election Misinformation specifically) to be 100% effective – nor would an approach that attempts to achieve such a result be reasonable or proportionate, as required by Article 35 DSA. Erring on the side of removal in an attempt to guarantee 100% effectiveness would very likely result in over-removal, which would be inconsistent with users' fundamental rights, including freedom of expression.

c. How does the provider of TikTok ensure that search suggestions across its platform direct users to authoritative sources?

38. TikTok understands that, in this ***question 3(c)***, the Commission is inquiring specifically about how TikTok ensures that its Search Suggestions, as defined in response to ***question 3(a)*** above, direct users to authoritative sources relating to electoral processes and civic discourse.

39. First, Search Suggestions are intended to direct users to helpful content. As explained in response to ***question 3(b)*** above, where TikTok identifies that a Search Suggestion matches with a verified account (which may include GPPPAs), TikTok displays the "verified badge" alongside it so users can see that by clicking on the Search Suggestion a verified account will be returned in the Search results.

40. However, otherwise, Search Suggestions are not linked to, and do not direct users to, specific sources. They simply suggest search terms for searches that a user may wish to carry out. The actual content and sources of information are presented to users in the Search *results*, i.e., what is displayed if the user does in fact choose to carry out a particular search. TikTok has implemented various mitigation measures in respect of Search *results* which help to identify and direct users to authoritative sources. To assist the Commission and provide the information that TikTok understands it is seeking with this question, TikTok provides the information below in relation to relevant measures it has implemented affecting Search results. See also TikTok's responses to ***question 2*** above (where TikTok discusses the interplay between content moderation and recommender systems); ***question 3(b)*** above (where TikTok explains, among other things, how it avoids Search Suggestions that it already knows are likely to return Election Misinformation); and ***question 4*** below (where TikTok explains how it implemented Search banners, discussed below, during the EU Elections).

41. TikTok has implemented the following mitigation measures which specifically help to ensure that, in the context of elections, Search results include content from authoritative sources:

    a. Election Centres and Search banners: As explained, e.g., in TikTok's Year 1 and Year 2 RARs, during elections, TikTok directs users to "Election Centres" through (among other entry points – see further TikTok's response to ***question 4*** below) 'labels' or 'banners' applied on Search results. For example, as explained in TikTok's Year 2 RAR[16], in relation to the EU Elections, from 28 March 2024, TikTok launched individual in-app "Election Centres" for each EU Member State, available (as relevant in each Member State) in 27 EU languages. TikTok's goal was to keep its communities informed about the EU Elections with trusted and authoritative information. Working with authoritative external stakeholders

[16] p. 115.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

(e.g., national electoral commissions, the European Parliament, and civic society organisations including fact-checkers) TikTok developed these Election Centres to connect its users with reliable voting information, including when, where, and how to vote – and, ultimately, with the election results themselves. As referred to below, all 27 EU Elections Election Centres contained a media literacy section that included videos from TikTok's trusted fact-checking partners and local media literacy bodies. Through these measures, TikTok emphasised that it supports official civic and electoral institutions and authorities.

b. Media literacy content and campaigns: As explained in TikTok's Year 2 RAR[17], starting from 2023 and throughout 2024, TikTok also introduced media literacy resources in the Election Centres (including in all the EU Election Centres), including videos from trusted fact-checking partners with, e.g., tips about how to develop critical thinking skills, identifying Election Misinformation, manipulated media and deep fakes.

d. How does TikTok apply content moderation on its search suggestion systems to mitigate risks for electoral processes and civic discourse, especially from users intentionally circumventing mitigation measures ([18])?

42. TikTok understands that, in this ***question 3(d)***, the Commission is inquiring specifically about the content moderation measures TikTok has implemented in relation to Search Suggestions, as defined in response to ***question 3(a)*** above, to mitigate risks for electoral processes and civic discourse. This information is provided in response to ***question 3(b)*** above.

43. TikTok reacted swiftly following publication of, and in response to, the specific report to which the Commission refers in its question[19] (the "**September ISD Report**"). In September 2024, TikTok internally escalated the terms flagged in the September ISD Report for immediate review and removal from Search Suggestions. TikTok is committed to continuous improvement and investment in developing and enhancing its variants detection model to boost its performance and coverage and TikTok's ability to proactively identify violative Search terms.

44. As part of its ongoing internal risk tracking and reporting processes, TikTok identified the risk of deliberately misspelled search terms being employed to evade TikTok's Search content moderation measures, and added new terms intended to 'instruct' the model to detect variations of violative search terms and the risk of users intentionally circumventing mitigation measures.

Further comments

45. TikTok makes the following further comments in relation to the Commission's question and the September ISD Report:

46. First, although the September ISD Report helpfully highlighted specific instances of violative content and Search Suggestions on the Platform, to which TikTok responded swiftly as noted above, that does not mean the balance of the September ISD Report is valid. For example, TikTok does not agree with its allegations or sweeping "*key findings*". Similar considerations apply to the September ISD Report as to those explained in relation to the Report and Blogpost in response to ***question 3(a)*** above, and TikTok respectfully submits that the September ISD Report do not contain such research or suggestions, and should not contribute to the Commission's assessment of TikTok's compliance with Articles 34 or 35 DSA.

47. Second, as in TikTok's response to ***question 3(a)*** above, TikTok refers to its comments in ***question 5 below*** in relation to Articles 8 and 14 DSA, and repeats them as applicable to this ***question 3***.

---

[17] p. 114.

[18] For example, see: https://www.isdglobal.org/wp-content/uploads/2024/09/TikTok-White-Supremacy.pdf, accessed on 26 September 2024.

[19] At https://www.isdglobal.org/wp-content/uploads/2024/09/TikTok-White-Supremacy.pdf.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

c. What is the accuracy of content moderation systems applied on TikTok's search suggestion systems across official EU languages, both in offline testing and online production environments?

48. As explained in response to ***question 3(b)*** above, TikTok has implemented both human and automated content moderation in respect of Search Suggestions. TikTok understands that, in this ***question 3(e)***, the Commission is inquiring specifically about TikTok's automated content moderation systems (i.e., those which may be subject to "*offline testing and online production environments*") applied on TikTok's Search Suggestions relating to electoral processes and civic discourse.

49. TikTok does not assess the accuracy of its automated content moderation systems in respect of Search Suggestions for each of the policies under its Community Guidelines individually, and therefore does not have data specific to electoral processes and civic discourse.

4. Ahead of the last European elections, the provider of TikTok announced the introduction of labels on elections-related content. Labels would direct people to relevant Election Centres ([20]). In a post-election blogpost, the provider of TikTok stated that *"We actively directed people to these Centres through prompts on relevant election content and election-related searches, which were labelled with notice tags and search banners respectively"* ([21]). Recent research from Finland ([22]) showed an inconsistent approach to labelling of elections-related content in the months before the European elections. In this regard, please provide the information requested below for the different systems, features or sections in the platform's online interface, including the 'For You Feed', 'You May Like', 'LIVE', suggested accounts, suggested search, displayed ads, and banners:

a. A detailed description and internal documents concerning how the labelling system related to elections-content is implemented by the provider of TikTok in the relevant period, including information on how recommendations are affected by these labels, as well as on any testing in place to assess the effectiveness of labelling as a mitigation measure;

50. TikTok understands that, in this ***question 4(a)***, the Commission is asking specifically about the labels that TikTok applied to:

    a. elections-related videos (irrespective of where in the Platform's online interface those videos were recommended) ("**notice tags**");

    b. elections-related Search results ("**search banners**"); and

    c. banners on the LIVE feed ("**LIVE room banners**"),

    in relation to the EU Elections, that would direct users to EU Elections-related Election Centres (described in response to ***question 3(c)*** above) (notice tags, search banners and LIVE room banners together, "**Election Labels**" or "**Election Labelling**").

51. Distinct from Election Labelling, TikTok also directed users to EU Elections-related Election Centres through push notifications that TikTok sent to users ("**Push Notifications**").

52. As detailed further below and in the documents enclosed in response to the Commission's request for relevant internal documents, Election Labelling was applied on top of / alongside videos and Search

[20] See: https://newsroom.tiktok.com/en-eu/our-work-to-prepare-for-the-2024-european-elections, accessed on 26 September 2024.

[21] See: https://newsroom.tiktok.com/en-eu/how-we-kept-our-community-safe-during-the-2024-europeanparliament-elections, accessed on 26 September 2024.

[22] See: https://crossover.social/wp-content/uploads/2024/07/TheTiktokEffect.pdf, accessed on 26 September 2024.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

results, and within the LIVE feed, and did not otherwise affect those recommendations from the feeds and features to which they were applied. Push Notifications also did not affect those recommendations.

53. With respect to notice tag Election Labelling (specifically) applied on videos:

    a. Notice tags were automatically displayed on: (i) new videos (i.e., uploaded within one month before the launch of the Election Centres on 28 March 2024) with hashtags matching those in TikTok's specially compiled EU Elections keywords list; and (ii) videos from GPPPAs. Notice tags were not applied to videos containing hidden captions.

    b. The notice tags were located in the "infobar" at the bottom of the user's screen and read "*Get info on the European Elections*". (For countries that also had national/local elections during the EU Elections period, the notice tag referred to those as well.)

    c. Notice tags were displayed to users in countries where the EU Elections were taking place in their local language, taking specific language settings into account. When the user clicked on the notice tag, they would be directed to their local in-app Election Centre.

    d. Notice tags were launched in the months ahead of, and ended after, the EU Elections in June 2024 after the elections..

    e. An English language example showing how notice tags were implemented is included as **Figure 1** below.

54. With respect to search banner Election Labelling (specifically), applied on Search results:

    a. Similar to notice tags, search banners were also displayed based on a set of keywords that TikTok identified ahead of the EU Elections, related to topics such as political parties, candidates, party slogans, etc. In compiling this keywords list, TikTok ensured that all major political parties had equal representation within the keywords list to avoid any potential bias for or against any particular parties or candidates.

    b. Search banners were displayed to users in countries where the EU Elections were taking place. The search banners were located at the top of Search results. Their heading read "*Election Awareness*" and their body text read "*Verify election information and voting guidance using authoritative sources*" (with an invitation to click through to the Election Centre and "*Learn more*"). As with notice tags, localised translations were displayed to users in different countries.

    c. Search banners were launched in the months ahead of, and ended after, the EU Elections on 9 June 2024.

    d. An English language example showing how search banners were implemented is included as **Figure 2** below.

55. With respect to LIVE room banner Election Labelling:

    a. LIVE room banners were displayed to users in countries where the EU Elections were taking place. LIVE room banners, which were displayed (in the relevant localised language) at the top of the "For You" tab (which is distinct from FYF) when users clicked "Explore" on the LIVE feed.

    b. LIVE room banners were launched in May 2024, and ended after the EU Elections, in June 2024.

    c. An English language example showing how LIVE room banners were implemented is included as **Figure 3** below.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013700

56. Additionally, TikTok sent two Push Notifications: (i) the first was sent within a week before the EU Elections in each market, and read "*Want to learn more about the European elections? Visit our Election Centre here.*"; and (ii) the second was sent on the EU Elections election day in each market, and read "*The European elections are taking place today. Find out more here*".

57. Election Labelling is one of a range of complementary mitigation measures, as detailed in TikTok's Year 1 and Year 2 RARs, that TikTok has implemented to address elections-related risks on the Platform. TikTok recorded certain metrics demonstrating the wide reach of TikTok's Election Labelling. For example, from 6 May to 9 June 2024:

    a. for notice tags, there were 4,753,349,948 views; and

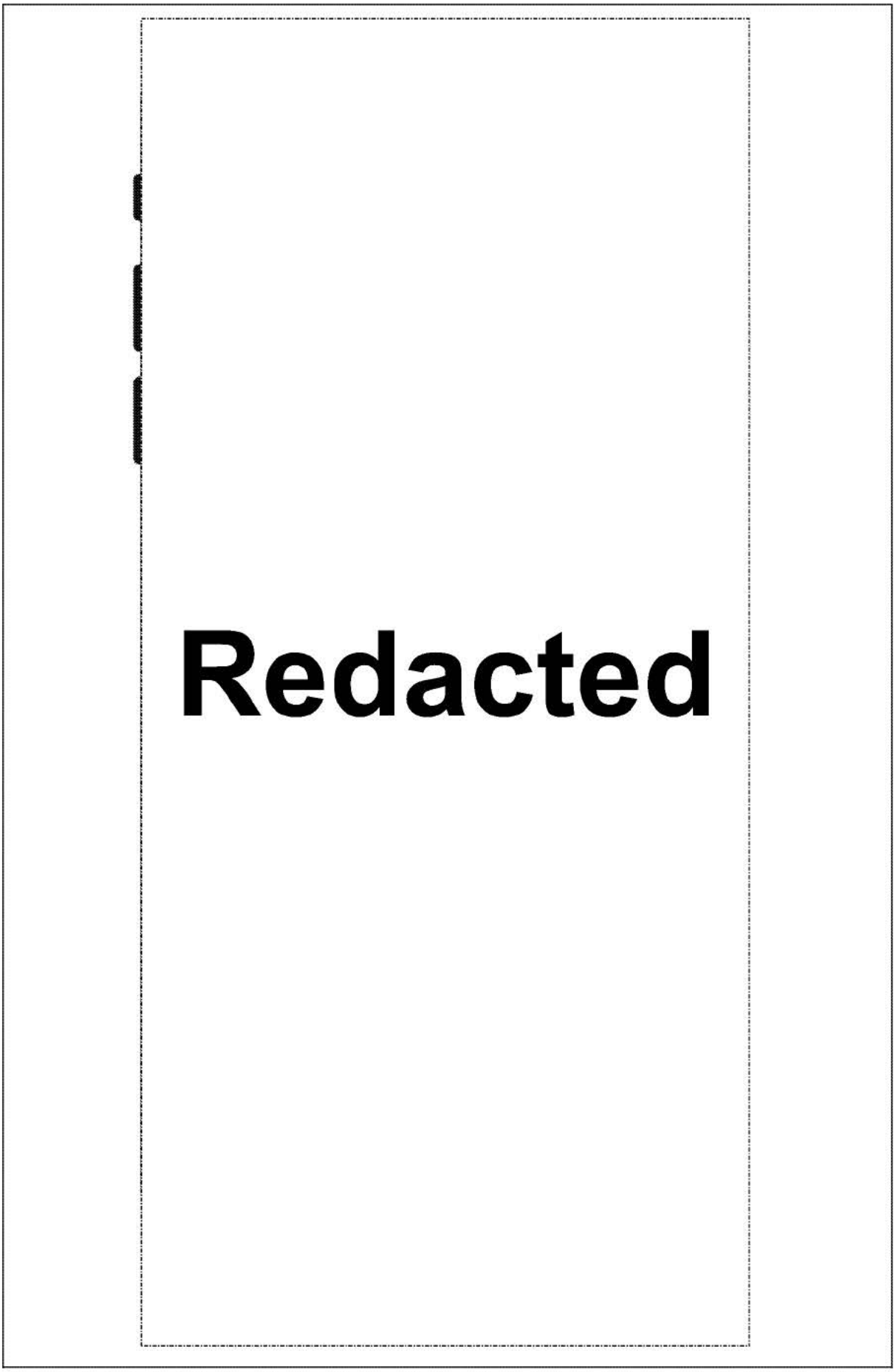
    b. for search banners, there were 63,613,989 views.

Further details are provided in TikTok's Year 2 RAR.[23]

[23] p. 67.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013701



CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

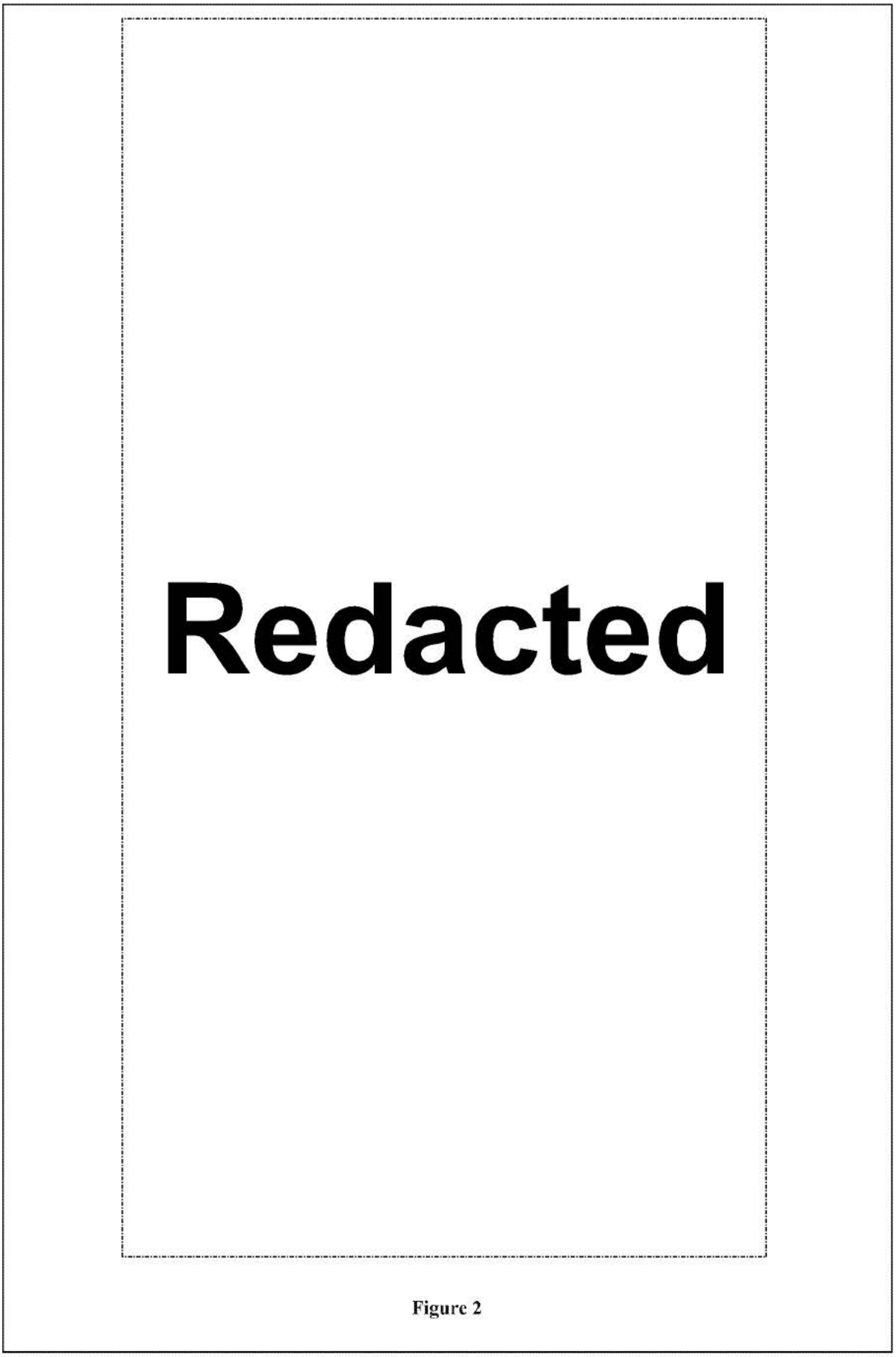


Figure 2

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Redacted

Figure 3

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013704

58. Copies of the requested documents TikTok has been able to identify in response to this ***question 4(a)*** are provided as Annexes 042 - 069.

b. Internal documents detailing guidelines, policies, and practices to review, revise, and adapt the previously referenced labelled content, including any performance metrics or internal controls currently tracked that may trigger changes in these guidelines, policies, and practices, such as false negatives, false positives, or user complaints, among others;

59. Copies of the requested documents TikTok has been able to identify in response to this ***question 4(b)*** are provided as Annexes 065 and 069.

c. Information on whether any differences exist across EU Member States and official EU languages, including on testing outcomes, labelling strategies and effectiveness, as well as on the tracked performance metrics or internal controls.

60. TikTok understands that, in this ***question 4(c)***, further to ***question 4(a)*** and TikTok's response to that question above, the Commission is asking about whether material differences exist between EU Member States or official EU languages with respect to TikTok's implementation of Election Labelling (as defined above) and the effectiveness of Election Labelling as a mitigation measure against elections-related risks on the Platform.

61. Please see TikTok's response to ***question 4(a)*** above. As explained in that response, Election Labelling (and Election Centres) was implemented across all EU Member States and official languages and there were no material differences between EU Member States or official languages with respect to how it was designed and implemented. However, voting during the EU Elections took place on a national level, according to different national procedures. Accordingly, although the EU-wide Election Centre resources were available to all EU users, they contained region-specific information, as they were created in partnership with local electoral commissions, the EU Parliament and civil society organisations and fact-checkers. All EU countries were regarded equally by TikTok in this regard, and no single market considered more 'important'. See also TikTok's Elections & Civic Integrity Module Questionnaire.[24]

62. TikTok's regional teams, who have a local understanding of the language and context, developed the list of keywords that TikTok used to identify content and searches on which to apply Election Labelling. As part of its post-EU Elections analysis, TikTok did record that, for certain countries, there were issues whereby some of the hashtags/keywords originally identified to trigger the notice tags resulted in excessive and/or insufficient application of the notice tags. However, TikTok had identified this issue quickly ahead of the EU Elections taking place and amended the list of hashtags/keywords accordingly.

[24] p. 25.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

5. In its 2023 risk assessment report, the provider of TikTok asserts that "*TikTok prohibits impersonation, which refers to accounts that pose as another person or entity in a deceptive manner. TikTok also reviews political accounts on the Platform in order to verify whether they are bona fide. TikTok does allow parody accounts*" ([25]). TikTok's Help Center states, "*TikTok is an entertainment platform where millions of people come every day to watch, share, and create entertaining, joyful content. Some of this content includes current events in elections and political issues. Our priority in this space is to make sure we keep harmful misinformation off the platform and ensure our community has a positive and welcoming experience. That's why we classify relevant political TikTok accounts as a Government, Politician, and Political Party Account (GPPPA) and apply a number of policies that help prevent misuse of certain features*" ([26]). Recent research indicates a significant number of highly active accounts spreading political content on the platform, while having neither clear affiliations nor indications of authenticity ([27]). Please provide:

a. A detailed description of the internal review process for political accounts mentioned in the 2023 "Risk Assessment 2: Integrity & Misinformation Module 1: Risks to Election & Civic Integrity" ([28]);

63. TikTok notes that the Commission's introduction to this ***question 5*** juxtaposes: (i) quotations referring to TikTok's prohibition on impersonation, and processes for identifying bona fide political accounts; with (ii) a suggestion that there are (nevertheless) "*a significant number of highly active accounts spreading political content on the platform, while having neither clear affiliations nor indications of authenticity*". While TikTok is, of course, required under Articles 34 and 35 DSA, respectively, to diligently assess systemic risks stemming from the design or functioning of its services – and to put in place reasonable, proportionate, and effective mitigation measures concerning such identified systemic risks – as the Commission will be aware: (i) Article 8 and Recital 30 DSA expressly state that the DSA does not impose, and nothing in it should be construed as imposing, a general monitoring obligation or a general active fact-finding obligation on providers, nor a general obligation for providers to take proactive measures in relation to illegal content; and (ii) Article 14(4) DSA requires providers to enforce restrictions in their terms and conditions in a "*diligent, objective and proportionate manner*" with due regard to, among other things, users' fundamental rights such as freedom of expression. TikTok understands that this question does not suggest otherwise.

64. The Commission's introduction to ***question 5*** also refers to three separate mitigation measures that TikTok has implemented:

    a. First, certain accounts belonging to political persons, entities or organisations (i.e., "political accounts") are classified as "Government, Politician, and Political Party Accounts" ("**GPPPAs**"), under the various different categories in TikTok's GPPPAs policy[29]. TikTok reviews political accounts to identify whether they are GPPPAs and, for accounts identified as such, applies an internal GPPPAs label in TikTok's systems (see further below). Although the Commission has referred to p. 36 of TikTok's Election & Civic Integrity Module Report, TikTok understands that, in this ***question 5(a)***, the Commission intends to ask about the internal review process by which TikTok identifies which political accounts are GPPPAs. TikTok explains this below.

---

[25] See: TikTok DSA risk assessment report, submitted by the provider of TikTok to the European Commission on 29 September 2023, registered under Ares (2023)6749619, p. 36.

[26] See: https://support.tiktok.com/en/using-tiktok/growing-your-audience/government-politician-and-political-party-accounts, accessed on 26 September 2024.

[27] See: https://democracy-reporting.org/en/office/EU/publications/inauthentic-behaviour-on-tiktok-concerning-accounts-support-the-afd-and-rassemblement-national, accessed on 26 September 2024.

[28] See: Risk Assessment 2: Integrity & Misinformation. Module 1: Risks to Election & Civic Integrity, registered under Ares(2023)8221551, p. 36.

[29] See further https://support.tiktok.com/en/using-tiktok/growing-your-audience/government-politician-and-political-party-accounts ("**TikTok's GPPPAs Help Centre article**").

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013706

b. Second, under TikTok's Integrity & Authenticity policies, TikTok prohibits impersonation and other inauthentic accounts. This includes deceptive impersonation of any person (including, but not limited to, political accounts that have been labelled as GPPPAs and other political accounts). The Commission has asked about, and TikTok explains, this further in ***questions 5(c)*** and ***5(d)*** below.

c. Third, under TikTok's verified accounts policy, TikTok awards a "verified badge" to certain notable, authentic accounts that TikTok has confirmed belong to the person (including, but not limited to, political accounts that have been labelled as GPPPAs and other political accounts) or brand they represent. Users can be sure that when they see the "verified badge" on an account, the verified account is exactly who they say they are, rather than an impersonation, parody or fan account. "Verified badges", and key factors informing TikTok's decision as to whether a particular account should be verified, are explained further in TikTok's Help Centre.[30]

65. As part of the Elections Integrity Programme, TikTok's local and global Government Relations teams review political accounts on the Platform and determine whether such accounts should be classified as GPPPAs in accordance with the GPPPA categories in TikTok's policy definition (available in TikTok's GPPPAs Help Centre article, and see the Elections & Civic Integrity Module Report[31]), to help ensure that GPPPA account-level restrictions (see further below) are applied to relevant accounts. The GPPPAs policy definition, and the guidelines for the definition, verification, and internal labelling of GPPPAs, were developed by a cross-functional working group from TikTok's Public Policy, PR and TnS teams.

66. This review of political accounts to determine if they are GPPPAs is currently undertaken by TikTok's Government Relations teams across the EU who use their local knowledge and market expertise to identify political accounts on TikTok that are potential GPPPAs – e.g., based on their own awareness of specific accounts, or by conducting searches. They then review those political accounts and confirm whether they fall within any of the GPPPA categories listed in TikTok's policy.

67. Where an account is identified as being a GPPPA, TikTok applies an "Institutional Type/Level" label to the account, internally in TikTok's systems, related to the GPPPA category that applies to the account. For example, there are labels for Government Entities, Political Individuals and Heads of State.

68. All GPPPAs identified and classified in accordance with TikTok's policy definition will be internally labelled as GPPPAs; it is not optional. When TikTok labels a political account as a GPPPA, it means that account will be subject to TikTok's GPPPA account restrictions (see further TikTok's response to ***question 5(f)*** below and TikTok's GPPPAs Help Centre article).

69. Once identified and classified as a GPPPA, GPPPAs are informed by TikTok's Government Relations teams of the separate possibility of going through the account verification process (i.e., to obtain a publicly visible "verified badge"). Account verification for GPPPAs and other political accounts within the EU is optional.

a. As explained in the Elections & Civic Integrity Module Report, through TikTok's Public Policy, Government Relations and Outreach & Partnerships Management ("**OPM**") teams, TikTok engages with Election Commissions/authorities to obtain authoritative information from them regarding the conduct of elections under their remit. To assist with identifying and classifying GPPPAs, particularly ahead of important elections, TikTok's Government Relations teams may also reach out to the main political parties, prominent candidates and other entities taking part in those elections. In addition to establishing clear points of contact between the GPPPA and TikTok, this outreach can also include education about

[30] See https://support.tiktok.[Redacted]ktok.

[31] pp. 40-41 and 69.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TikTok and its policies, including information about the restrictions that apply to GPPPAs, and how they can become verified.

b. For example, in relation to the EU Elections, TikTok's teams actively engaged with all European political groups and political parties and ensured that as many EU GPPPAs as possible were identified and labelled correctly ahead of the elections. European leaders, ministers, and political parties join TikTok on a regular basis – indeed, many of them did ahead of the EU Elections, including approximately 30% of the sitting Members of European Parliament ("**MEPs**") and the EU Parliament itself. These politicians and parties play an important role on the Platform – TikTok believes that "verified" accounts belonging to politicians and institutions provide the electorate with another route to access their representatives, and additional trusted voices in the shared fight against misinformation. Before the EU Elections, TikTok engaged with both sitting MEPs and candidates from across the EU Member States to ensure that those using TikTok had necessary information to do so safely and authentically.

70. After an election cycle in a particular country, the local list of GPPPAs is updated by TikTok's Public Policy team. The relevant internal GPPPA labels are then applied/updated accordingly.

71. GPPPA labels can also be reviewed at the request of a GPPPA holder who believes that their account has been mislabelled as such. Requests can be sent via the TikTok app or directly to TikTok's Government Relations or Public Policy teams.

b. A detailed description of the measures in place to detect political accounts on TikTok, including performance metrics or internal controls used to evaluate the detection effectiveness (e.g., precision, recall, false negatives);

72. TikTok understands that, in this ***question 5(b)***, the Commission is asking whether TikTok has in place automated measures to detect political accounts that may qualify as GPPPAs.

73. At present, TikTok does not have any such automated measures (or, therefore, performance metrics or internal controls for evaluating their effectiveness). As explained in TikTok's response to ***question 5(a)*** above, the identification of political accounts and review as to whether they are GPPPAs is undertaken by the Government Relations team and is a manual process.

c. A detailed description of the measures in place to detect inauthentic accounts or impersonation on TikTok, particularly for political accounts, including performance metrics or internal controls used to evaluate the detection effectiveness (e.g., precision, recall, false negatives);

74. TikTok understands that, in this ***question 5(c)***, the Commission is asking whether TikTok has any automated systems and measures to detect inauthentic or impersonating political accounts. In this regard, please see the comments in TikTok's response to ***question 5(a)*** above regarding Arts. 8 and 14 DSA and the premise of the Commission's question.

75. TikTok operates a number of measures, including automated measures, to identify and remove or restrict access to violative content on the Platform, including inauthentic and impersonation accounts – regardless of whether those concern a political account or not. However, as explained in response to ***questions 5(a)*** and ***5(b)*** above, TikTok does not have any automated measures in place to detect political accounts.

76. TikTok's Integrity & Authenticity policies, which address Spam and Deceptive Account Behaviours[32], expressly prohibit account behaviour that may deceive, mislead or spam TikTok's community. TikTok's assessment of whether an account (as opposed to, for example, content or behaviour) is

[32] See https://www.tiktok.com/community-guidelines/en/integrity-authenticity#6.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

"inauthentic" takes into consideration different criteria depending on the context. On the Platform, "inauthentic" accounts can include:

a. spam accounts (e.g., accounts that are operated in bulk, through unauthorised automation, or in order to distribute high-volume commercial content, or networks of accounts that represent similar entities or post similar content to lead others to specific locations, such as other accounts, websites, and businesses);

b. accounts related to a covert influence operation ("**CIO**") (coordinated, inauthentic behaviours where networks of accounts work together to mislead people or TikTok's systems and try to strategically influence public discussion, such as attempting to undermine the results of an election); and

c. impersonation accounts (accounts that misleadingly or deceptively pose as another real person or entity, such as using someone's name, biographical details, content, or image without disclosing it, or presenting as a person or entity that does not exist (a fake persona) with a demonstrated intent to mislead others on the Platform).

77. There can be overlap between these categories of inauthentic accounts. For example, impersonation may be accomplished through automated means, and spam/bot accounts are usually automated with scripts or emulators – and TikTok has identified and taken action in respect of multiple clusters involving thousands of automated impersonation accounts.

78. TikTok's automated measures to identify inauthentic accounts include:

a. proprietary detection models to proactively monitor and detect (attempted) automated creation of inauthentic/spam accounts, based on a number of technical signals related to manner of user registration, account traffic patterns, and engagement and content. This includes the use of automation to register or operate accounts in bulk, to artificially increase engagement signals, or to circumvent enforcement of TikTok's policies. TikTok's automated detection models remove inauthentic registered accounts, based on these signals, if certain thresholds are met; and

b. a partially automated process using an automated tool to conduct "impersonation sweeps". Especially during elections or other such high-risk events, if TikTok suspects certain genuine accounts (e.g., a GPPPA) might be being impersonated, TikTok can use the tool to sweep the Platform, looking at factors such as textual and visual similarities across the usernames and avatars[33] on other accounts. Accounts that the tool identifies as being above a certain percentage match with the potentially impersonated account are then queued for manual review.

79. In addition to the automated systems measures which TikTok understands the Commission's question to concern, TikTok notes that, as detailed in the Elections & Civic Integrity Module Report[34], TikTok also identifies impersonation and other inauthentic accounts through manual processes, including:

a. manual impersonation sweeps, which are similar to the automated sweeps referred to above, but which can be more targeted and take into account a wider range of factors, but across a smaller volume of potentially impersonating accounts; and

b. manually monitoring reports of inauthentic accounts received from individual users and trusted flaggers, in order to detect larger clusters or similar inauthentic behaviours; and

c. investigations to detect and disrupt CIOs, including by removing associated accounts from the Platform. CIO investigations are resource-intensive and require in-depth analysis to ensure sufficiently high confidence in proposed actions. Such investigations focus on behaviour and assessing linkages between accounts and techniques to determine if actors

[33] See further https://newsroom.tiktok.com/en-us/express-yourself-through-tiktok-avatars.
[34] pp. 33-37.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013709

are engaging in a coordinated effort to mislead TikTok's systems or its community. TikTok uses several types of information (including open-source and proprietary) to assess CIOs, and leverages a standard framework of confidence assessment to help ensure that it makes consistent and accurate determinations. Where TikTok's teams have the necessary high degree of confidence that an account is engaged in a covert influence operation or connected to networks previously removed as part of a CIO, it is removed from the Platform.

80. As reported in TikTok's September 2024 COPD Report, metrics indicating the (successful) performance of these measures include that, in the first half of 2024, TikTok:

    a. removed approximately 42 million inauthentic accounts or botnets (which can include automated, partially automated, or non-automated accounts) from the Platform throughout the EU; and

    b. banned over 12,000 accounts throughout the EU under TikTok's impersonation policy.

81. Further, as reported in TikTok's Community Guidelines Enforcement Report for the quarter April - June 2024:

    a. In respect of videos posted to the Platform which violated TikTok's Integrity & Authenticity policy related to fake engagement, TikTok had a:

        i. proactive removal rate (i.e., removed before receipt of a user report) of 94.4%;

        ii. a removal rate 'before any views' of 45.7%; and

        iii. a removal rate 'within 24 hours' of the video being uploaded to the Platform of 79.6%; and

    b. TikTok prevented the creation of 347,888,700 fake accounts

Further comments

82. As part of TikTok's response to the Commission's question about measures to detect inauthentic political accounts, TikTok notes the Commission's reference to (and implied reliance on) the article by Democracy Reporting International ("**DRI**") (the "**DRI Article**"). As with the Report and Blogpost referred to in ***question 3***, TikTok does not propose to respond to the DRI Article comprehensively here. Nonetheless, TikTok wishes to note that it does not accept the DRI Article's suggestions concerning its compliance with Articles 34 and 35 DSA. In particular, TikTok does not agree with the DRI Article's:

    a. apparent position that an unofficial account or an account with "*unclear affiliation*" is necessarily "inauthentic". In TikTok's view, such accounts would only be inauthentic if they are also deceptive or misleading (see further the response to ***question 5(d)*** below);

    b. (unexplained) allegations that TikTok has not complied with its obligations under Articles 34 and 35 DSA; or

    c. suggestion that the presence of *any* (allegedly) inauthentic accounts on the Platform indicates that TikTok has not complied with its obligations under the DSA. As explained in TikTok's Elections & Civic Integrity Module Report[35], content moderation and removal of Election Misinformation will never be 100% effective, nor would an approach that attempts to achieve such a result regardless of the risk of over-removal be reasonable, proportionate or consistent with users' fundamental rights, as required by Art. 35 DSA.

[35] pp. 20 and 60.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013710

83. More generally, please see TikTok's response to RFI Ref. Ares(2024)8523013 (the "**First Romanian Elections RFI**"), which discusses the difficult balance between, on the one hand, tackling Election Misinformation and, on the other hand, protecting users' fundamental rights and freedom of expression, which is particularly important in the context of elections. The potential negative impacts on freedom of expression implied by some of the proposals contained in the DRI Article would be concerning and significant.

d. A detailed description of how the provider of TikTok distinguishes between inauthentic accounts and parody accounts;

84. TikTok understands that, in this ***question 5(d)***, the Commission is asking how TikTok distinguishes between parody or fan-based accounts (which are permitted on its Platform) and impersonation accounts (which are not). As set out in TikTok's Integrity and Authenticity Community Guidelines, the former are allowed, as long as the account name clearly discloses that it is a parody or fan-based account (note that account name is different from the @username). By contrast, impersonation accounts include accounts that: (i) pose as another real person or entity without disclosing that they are a fan or parody account in the account name, such as using someone's name, biographical details, content, or image without disclosing it; or (ii) present as a person or entity that does not exist (a fake persona) with a demonstrated intent to mislead others on the Platform.

85. As set out in paragraph 78 above, TikTok employs automated measures to identify inauthentic accounts. If these processes surface a potentially inauthentic account, TikTok will only treat it as a fan or parody account if the account name clearly discloses it as such. If it does not contain such a disclosure, TikTok will consider if it appears to be inauthentic. Signals that indicate that an account-holder is posing as another person or entity include, among other things, using someone else's name, biographical information, or profile picture, and/or making false representations in video content. Naturally, it can be difficult for automated measures to evaluate such nuanced signals. Therefore, TikTok also identifies inauthentic accounts through manual processes, as set out in paragraph 78 above. Further, even based on a manual review, it can be challenging for an intermediary like TikTok to: (i) identify potentially inauthentic accounts; and (ii) confirm whether they are, in fact, inauthentic.

e. For Government, Politician, and Political Party Accounts (GPPPAs), clarify whether any recommender system rules, configurations or parameters affect their visibility on the platform;

86. There are no such recommender system rules, configurations or parameters on the Platform. Please see further paragraph 36 of TikTok's response to ***Section 1.1*** of RFI Ref.Ares(2024)8716531 (the "**Second Romanian Elections RFI**").

87. With respect to more general account-level restrictions in place for GPPPAs, please see TikTok's response to ***question 5(f)*** below and TikTok's GPPPAs Help Centre article.[36]

88. TikTok's response to ***question 11*** below explains TikTok's "strikes"-based enforcement approach against accounts which repeatedly violate its rules. However, as explained in that response, TikTok enforces different account restrictions against "**public interest accounts**" (i.e., GPPPAs and news entities), because of the role they play in civic processes and civil society, in keeping with TikTok's commitment to human rights and freedom of expression. If such accounts reach the strike limit set for all accounts, instead of being banned, the account will be temporarily ineligible to appear in the FYF and Following feed for 90 days (see TikTok's response to ***question 10(a)*** below regarding the effect of FYF Eligibility Standards).

[36] See https://support.tiktok.com/en/using-tiktok/growing-your-audience/government-politician-and-political-party-accounts.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

f. Please explain all the measures in place for GPPPAs to mitigate any risk to electoral processes and civic discourse stemming from its specific recommender systems;

89. TikTok understands that, in this ***question 5(f)***, the Commission is asking about any measures TikTok has in place that apply specifically to GPPPAs, to mitigate any risks to electoral processes and civic discourse stemming from recommendation by TikTok's recommender systems of content provided by GPPPAs.

90. As explained in response to ***question 5(e)*** above, there are no recommender system rules, configurations or parameters that affect the visibility of all GPPPAs (or content they provide) on the Platform, as a result of being labelled as (or provided by) a GPPPA – although certain exceptions to TikTok's regular enforcement policies apply to them (as explained further in response to ***question 11*** below).

91. However, as explained in TikTok's Elections & Civic Integrity Module Report[37] and GPPPAs Help Centre article,[38] TikTok has implemented certain more general measures relating to GPPPAs, which help to mitigate risks to electoral processes and civic discourse on the Platform.

92. As explained in response to ***question 5(a)*** above, all accounts identified as GPPPAs are internally labelled as such. Once an account has been labelled internally as a GPPPA, a number of policies are applied that help to prevent the risks of misuse of certain features. TikTok considers that a key way to mitigate such risks comes from circumventing certain activities, and accordingly GPPPAs' monetisation and advertising features are disabled (except for certain limited exceptions explained in TikTok's GPPPAs Help Centre article, such as to promote public health initiatives like COVID-19 booster programs). Solicitations for campaign fundraising by GPPPAs are also not permitted on the Platform. This includes content like a video from a politician asking for donations, or a political party directing people to a donation page on their website.

93. TikTok considers that restricting political advertising features is a particularly important mitigation measure for risks to electoral processes and civic discourse. As explained in TikTok's Year 2 RAR[39] and Elections & Civic Integrity Module Report,[40] TikTok therefore prohibits, at both a content and account level: (i) political figures (including GPPPAs, except for the limited circumstances outlined in paragraph 92 above) from promoting ads; and (ii) ads from referring to any political content, regardless of who is promoting them. As a further level of mitigation (e.g. for advertisers that have not been labelled as GPPPAs), TikTok has additional mechanisms at the point of ad creation to identify both account- and content-level violations, such that: (i) the ad can be rejected before it is posted; (ii) the ad can be removed after having been posted; and/or (iii) the advertiser's account can be suspended. More information about TikTok's political advertising policy can be found in TikTok's Political Advertising Policy[41] and GPPPA Business Help Centre Article.[42] See further TikTok's response to ***question 8*** of the First Romanian Elections RFI.

94. Further, to ensure accurate content moderation (e.g., to avoid false positives and false negatives), during the EU Elections, over 3,400 GPPPAs were subject to a proactive safeguarding process whereby reported content from their accounts was subject to review by multiple internal teams. This reflects TikTok's commitment to maintaining content integrity for GPPPAs, particularly in high-risk contexts such as elections.

---

[37] pp. 40-42.
[38] See https://support.tiktok.com/en/using-tiktok/growing-your-audience/government-politician-and-political-party-accounts.
[39] p. 64.
[40] p. 42.
[41] See https://ads.tiktok.com/help/article/tiktok-ads-policy-politics-religion-and-culture.
[42] See https://support.tiktok.com/en/using-tiktok/growing-your-audience/government-politician-and-political-party-accounts.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

6. In its risk assessment reports, the provider of TikTok mentions 'borderline content' or 'borderline claims' as part of the Fact-Checking Programme, i.e. content that does not violate TikTok's community guidelines, but it is close to the threshold required to be considered as violative. In the context of 'Election Misinformation', please provide:

a. A detailed explanation of how the specific recommender systems deal with 'borderline content' related to 'Election Misinformation'. Include information on any changes to the ranking, visibility, or prevalence of such content, such as after detection or after a certain threshold of content is reached;

95. TikTok's Fact-Checking Programme is described in Annex 5 of TikTok's Year 2 RAR and the response to ***question 11(b)*** below. The core objective / mission of the Fact-Checking Programme is to leverage the expertise of external fact-checking organisations to verify potential Election Misinformation (and other misinformation) claims, and to enable TikTok's Trust & Safety teams to determine appropriate enforcement actions. The Programme is grounded in TikTok's belief that technology companies should not alone make decisions about the veracity of borderline claims. Accordingly, TikTok partners with third-party fact-checking organisations as a critical component of its strategy to effectively tackle Election Misinformation (and other harmful misinformation).

96. The Fact-Checking Programme is integrated into a specialised misinformation moderation workflow in order to inform the moderation team's decisions on whether content violates the Election Misinformation (or other misinformation) policies. In the context of Election Misinformation, "*borderline*" content and claims do not refer to content that "*does not violate TikTok's community guidelines, but* [is] *close to the threshold required to be considered as violative*" as the Commission suggests in its question.

97. Rather, as explained further below, "*borderline*" Election Misinformation claims/content means content that:

    a. claims (impliedly or expressly) certain information to be true; but

    b. the veracity of that information cannot be determined by either TikTok's automated or human moderators (e.g., by reference to TikTok's repository of previously fact-checked claims – see below);

    c. the claim/content is not clearly otherwise violative of TikTok's Community Guidelines; and

    d. therefore an assessment and determination of the content/claim's veracity by specialist, third-party fact checkers is required, through TikTok's Fact-Checking Programme, which is explained further below.

98. The challenge with such content is that TikTok's automated systems and human moderators are unable to verify whether the information is factually incorrect or misleading.

99. The impact of the fact-checking process on the ranking, visibility, or prevalence of such content is as follows:

    a. While the content is being fact-checked, TikTok makes it ineligible for the FYF. FYF-ineligible content is not recommended on FYF and may be harder to find in Search.

    b. Once fact-checkers have assessed the content, TikTok's moderators apply the Community Guidelines accordingly:

        i. If the content is determined to be accurate (i.e., non-violative of TikTok's Election Misinformation policy), it remains on the Platform and is again eligible for recommendation on FYF (provided that it does not violate any other Community Guidelines).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013713

ii. If the content is confirmed as being Election Misinformation (or other misinformation that violates the Community Guidelines), TikTok's moderators apply TikTok's moderation policies accordingly.

iii. Where the fact checks are inconclusive (i.e., the independent fact-checkers cannot, or cannot yet, determine whether or not the content is Election Misinformation), TikTok refers to the content as "Unverified Election Claims" (which generally means the content's veracity cannot be confirmed – e.g., during unfolding events or emergencies), and will apply an 'unverified content label' to the video, making it ineligible for FYF to restrict its reach.

100. Where potentially violative content is identified by TikTok's automated technology but it is not clear-cut Election Misinformation (and therefore cannot be automatically removed with confidence), it is flagged for further review by human moderation teams. As explained in Annexes 3 and 5 of TikTok's Year 2 RAR, these content moderation teams include experienced and trained specialists. TikTok has dedicated teams and processes for its Integrity & Authenticity policies such as those covering Election Misinformation, including providing human content moderators with extensive and in-depth training on TikTok's misinformation policies, as well as enhanced tooling.

101. When potential Election Misinformation is detected and flagged for their review, these specialist Integrity & Authenticity moderators determine whether the content is accurate, false, or unverifiable. To do so, they: (i) first, consult a global database of claims that have previously been fact-checked by specialist fact-checkers; and (ii) second, route any new, evolving, or borderline claims (i.e., those which they cannot determine for themselves) to TikTok's Fact-Checking Programme for independent evaluation by a fact-checking partner with expertise in the market who will assess the video, conduct reporting, and verify whether the information is false.

102. For more information about TikTok's fact-checking partners, see https://www.tiktok.com/transparency/en-us/combating-misinformation/#GFCP and the response to ***question 11(b)*** below.)

b. Specify whether any guidelines, policies, and practices are in place for the detection and monitoring of 'borderline content' related to 'Election Misinformation'. Include any performance metrics or internal controls used to track the increase, decrease, or general evolution of 'borderline content'. For all provided performance metrics or internal controls, provide the results of tests carried out in both offline and online production environments.

103. Please see TikTok's response to ***question 6(a)*** above. As explained therein, "*borderline*" Election Misinformation means content that: (i) claims (impliedly or expressly) certain information to be true; but (ii) the veracity of that information cannot be determined by either TikTok's automated or human moderators; (iii) the claim/content is not clearly otherwise violative of TikTok's Community Guidelines; and (iv) therefore an assessment and determination of the content/claim's veracity by specialist, third-party fact checkers is required, through TikTok's Fact-Checking Programme, which is explained further below.

104. TikTok does not specifically track the increase, decrease, or general evolution of "*borderline*" content related to Election Misinformation.

105. Recent metrics related to the enforcement of TikTok's Integrity & Authenticity Community Guidelines (more generally) can be found in TikTok's September 2024 COPD Report (https://disinfocode.eu/signatory-report/tiktok-september-2024-report/) and Community Guidelines Enforcement Reports (https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2024-9).

106. Further, as explained in the Elections & Civic Integrity Module Questionnaire, TikTok's automated moderation models' precision and recall are measured from time to time and based on their interaction with human content moderation teams, assessing moderator "action rates" against content identified as

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

potentially violative by the models. The overall prevalence of violative content on the Platform is also measured regularly through the review and labelling of samples of posted content.

7. In its 2023 risk assessment, the provider of TikTok presents "*diversified content recommendations*" as a mitigation measure to prevent 'Election Misinformation' from being recommended or surfaced in a user's 'For You Feed': *"our recommendation system works to intersperse recommendations that might fall outside people's expressed preferences, offering an opportunity to discover new categories of content"* ([43]). Please provide with regards to *"political content"*:

a. Any internal guidelines, policies, and practices that may affect recommended content in terms of diversity or dispersion and any internal documents about testing performed *"to avoid recommending a series of similar content to protect against viewing too much of a content category that may be fine as a single video but potentially problematic if viewed in clusters"*[44];

107. From the Commission's reference to the explanation in TikTok's Elections & Civic Integrity Module Report that TikTok's mitigation measure to prevent Election Misinformation from being recommended in users' FYF include diversified content recommendations, TikTok understands that the "*political content*" about which the Commission is asking in this ***question 7*** means Election Misinformation that may appear in users' FYF.

108. Under TikTok's Community Guidelines, Election Misinformation is not allowed on the Platform. Videos assessed through TikTok's automated and/or human content moderation processes to violate TikTok's Community Guidelines, including the Misinformation or Civic & Election Integrity policies under the Integrity & Authenticity Community Guidelines, will be removed from the Platform. As reported in TikTok's April - June 2024 Community Guidelines Enforcement Report, approximately 98% of content that violates those policies is proactively removed. Over 96% of content violating the Civic & Election Integrity rules is removed before it receives any views at all. Therefore, TikTok does not have any guidelines, policies or practices in respect of diversity or dispersion of Election Misinformation in particular, or carry out any testing on that basis.

109. This is in contrast to content that is considered to relate to borderline extreme dieting and fitness, negative affect, borderline adult nudity and sexual activity, and mental health narratives, which are permitted on the Platform but may be problematic if viewed repeatedly in high quantities. As the Commission is aware from TikTok's response to the Youth Investigation RFI, the recommendation system underlying FYF does use dispersion techniques specifically for such content.

110. In respect of diversification, TikTok uses recommendations to offer its users a personalised experience and help them to discover interesting and relevant content among the wide variety of content on the Platform. These recommendations of suggested content are based on users' preferences, in particular those expressed through their interactions on TikTok, such as following an account or liking a post (see TikTok's response to ***question 1(d)*** above). However, TikTok recognises that too much of any particular type of content does not fit with the diverse discovery experience TikTok aims to create. To keep users' FYF fresh – and to interrupt repetitive patterns – the recommender system that powers FYF intersperses diverse types of content into the types of content that people already know they are interested in, offering them an opportunity to discover new categories of content that might fall outside their expressed preferences. In particular, FYF will not recommend two videos in a row made by the same creator or with the same sound. As well as enriching users' viewing experience, this can help to promote exposure to a range of ideas and perspectives on the Platform. TikTok notes that similar diversification measures also apply in respect of Friends feed, Following feed and LIVE.[45]

[43] See: Risk Assessment 2: Integrity & Misinformation. Module 1: Risks to Election & Civic Integrity, p. 27, registered under Ares(2023)8221551.

[44] See: Risk Assessment 2: Integrity & Misinformation. Module 1: Risks to Election & Civic Integrity, p. 28, registered under Ares(2023)8221551.

[45] See https://support.tiktok.com/en/using-tiktok/exploring-videos/how-tiktok-recommends-content.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

b. Detailed information on how such practices are assessed for effectiveness, specifying whether any additional measures are implemented in proximity to electoral processes, including information on how early they are activated.

111. As explained in response to ***question 7(a)*** above, TikTok does not have any specifically defined practices in respect of diversity or dispersion of Election Misinformation in particular, and does not assess its general diversification measures on that basis.

8. In its 2024 risk assessment, the provider of TikTok states *"TikTok has deployed improved machine learning models to facilitate the effective detection of potential Election Misinformation. The primary detection model now takes into account relevant subpolicies for more accurate identification of violative content (e.g. Election Misinformation, and Medical Misinformation). This means that the model now takes account of different possible types of misinformation and their respective varying characteristics, rather than purely basing its assessment on a binary outcome"* ([46]). With regards to risks to electoral processes and civic discourse, please provide the following information:

a. An exhaustive list of input factors, features, signals, information, and metadata used by automated models to detect content that may contain 'Election Misinformation'. Include an explanation of how the provided list of input factors is modified to detect 'Election Misinformation' in a timely manner;

112. TikTok understands that, in ***questions 8(a)*** to ***8(c)***, the Commission is asking about TikTok's automated models to detect Election Misinformation in video content uploaded by users to the Platform. TikTok has developed many such automated models in its efforts to continuously improve its detection and enforcement of the Community Guidelines and to update the logic underlying its content moderation models. TikTok understands that the automated models relevant to the Commission's question are: (i) TikTok's automated model which is trained to detect different types of misinformation (including medical misinformation and Election Misinformation) (the "**General Misinformation Model**"); and (ii) TikTok's automated model which is trained to detect Election Misinformation (the "**Election Misinformation Model**"; together with the General Misinformation Model, the "**Election Misinformation Models**"). Recent updates and improvements that TikTok has made to its Election Misinformation Models are explained in the Year 2 RAR and in the Elections & Civic Integrity Module Questionnaire.

113. Machine learning models can use many pieces of information to produce their outputs. TikTok therefore understands that, similarly to ***question 14*** of the Safety of Minors RFI, the Commission is seeking, in this ***question 8(a)***, to understand the *categories* of information principally used by TikTok's Election Misinformation Models, and how those are updated to efficiently detect Election Misinformation.

114. Given the volume of content posted to the Platform on a continuous basis (i.e., the millions of videos posted every day), all videos uploaded to TikTok are initially reviewed, in real-time, by TikTok's automated moderation technology. This aims to identify any content that violates the Community Guidelines.

115. TikTok's Election Misinformation Models use the following inputs to identify whether content contains Election Misinformation:

    a. country code (i.e., where the video was uploaded from);

    b. the video's title (including any hashtags);

[46] See: "*TikTok DSA risk assessment report 2024*" submitted by the provider of TikTok to the European Commission on 29 August 2024, registered under Ares(2024)6149012, p. 66.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013716

c. machine-readable transcripts of audio or music within a video (created using automatic speech recognition (ASR)); and

d. machine-readable transcripts of the visible text within the video (created using optical character recognition (OCR)), e.g., any sticker text.

116. The Election Misinformation Models compare the above information from the video against information from previously identified examples of Election Misinformation and fact-checked claims (see further TikTok's response to ***question 6*** above), in order to assess whether a particular statement or claim is Election Misinformation. On that basis they produce a "**confidence score**" (reflecting the assessed likelihood of the content containing Election Misinformation); (iii) depending on the confidence score, appropriate moderation action is taken (see ***question 8(b)*** below). Further explanation about how the information is assessed is provided in response to ***question 8(b)*** below.

117. Whereas the Election Misinformation Model is trained only on examples of Election Misinformation, and is a 'specialised' model, the General Misinformation Model is trained on examples of other misinformation as well. The General Misinformation Model can therefore take into account different types of misinformation and their respective varying characteristics.

118. The frequency with which TikTok retrains its Election Misinformation Models varies. Depending on the assessments made by TikTok's specialist teams, retraining typically occurs more frequently just before an election takes place, and, in any case, on a regular basis. However, a particular challenge with Election Misinformation is that it is often nuanced and election- and context-specific. For example, issues, candidates and information on how to vote vary from election to election and market to market. This can make Election Misinformation especially difficult to detect using Election Misinformation Models. In TikTok's experience, new, nuanced Election Misinformation claims can arise even the day after the Election Misinformation Models are retrained. In such cases, TikTok has other proactive strategies to reasonably and proportionally address new misinformation claims, including:

a. *Internal escalations*: TikTok has internal teams responsible for proactively identifying potential harms that are likely to manifest as user-generated content on the Platform. They do this in several ways – e.g., by conducting media sweeps for recent and emerging trends which might appear on the Platform;

b. *Requests from external partners*: Including government entities, onboarded Community Partnership Channel partners, or through measures set up under TikTok's Election Integrity Programme, through which various EU-based (and international) NGOs can report harmful content using a dedicated email channel or web portal. In this regard, TikTok's fact-checking partners are authorised to proactively identify content that may constitute harmful misinformation on the Platform and to suggest prominent misinformation that is circulating on other social media platforms or websites that may benefit from verification. TikTok also actively participated in the Code of Practice on Disinformation Rapid Response System (RRS) during several elections in 2024, to streamline the exchange of information among civil society organisations, fact-checkers, and online platforms. The RRS was a timebound cooperation framework, which allowed non-platform signatories to flag time sensitive content, accounts or trends that may present threats to the integrity of the electoral process; and

c. *Targeted proactive 'sweeps' of specific types of content*: For example, this may involve a targeted sweep focused on use of a specific keyword, URL, or hashtag identified as being associated with high risk content, such as certain 'viral' or popular content.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013717

b. A description of how the importance of the listed input factors is measured and how it impacts a content's position, ranking, or order within the displayed recommendations;

119. As explained in response to ***question 2*** above, content moderation systems are separate and distinct from recommender systems. The input factors taken into account by TikTok's Election Misinformation Models, listed in response to ***question 8(a)*** above, therefore do not impact the recommendations the recommender engine makes. As for content moderation decisions:

    a. If no violation is identified, the content will be available to view on the Platform.

    b. If a potential violation is found, the automated moderation system will either pass the content to TikTok's Safety teams for further review or remove it automatically if there is a high degree of confidence that the content violates the Community Guidelines.

    c. Such content may also breach the FYF Eligibility Standards (see https://www.tiktok.com/community-guidelines/en/fyf-standards). FYF-ineligible content will not be recommended through FYF, and may be harder to find in Search.

120. There is no pre-determined weighting of the particular inputs listed in response to ***question 8(a)*** above. They are simply different inputs that are taken into account.

c. A detailed description of when and how these machine learning models are used to detect potential 'Election Misinformation', specifying whether they are always active or only under certain circumstances, as well as on which type of content and platform surface they are used to detect potential 'Election Misinformation'.

121. Further to TikTok's responses to ***questions 8(a)*** and ***8(b)*** above, the Election Misinformation Models operate to detect Election Misinformation in video content uploaded by users (which can be viewed, e.g., in FYF, Following feed or Friends feed), not on a specific feed or feature.

122. The General Election Misinformation Model is continually active across all EU Member States.

123. The Election Misinformation Model is currently most effective at identifying Election Misinformation in English language content. It is currently active only in Ireland and Germany.

124. TikTok also notes that, as explained in TikTok's Year 2 RAR, all videos uploaded to the Platform, at all times, are subject to content moderation by TikTok's automated models and/or human moderators, at the point of upload. Only once a video has been through that content moderation process will it be available for potential recommendation to other users in FYF, Following, Friends or other Platform feeds. Videos assessed to violate TikTok's Community Guidelines, including the Misinformation or Civic & Election Integrity policies under the Integrity & Authenticity Community Guidelines, will be removed from the Platform. As reported in TikTok's April - June 2024 Community Guidelines Enforcement Report, approximately 98% of content that violates those policies is proactively removed. Over 96% of content violating the Civic & Election Integrity rules is removed before it receives any views at all.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013718

9. Research indicates that TikTok users are exposed to conspiracy theories, illegal content, misinformation and disinformation related to electoral processes and civic discourse ([47]). In its risk assessment reports, the provider of TikTok mentions "*manual review*" of content as a mitigation measure against the risk of "*virality*" or the risks associated with "*viral content*". In its 2023 risk assessment report, the provider of TikTok states that "[a]*s video content rises in popularity in terms of the number of video views, it will be flagged for further manual review*" ([48]). Similarly, in its 2024 risk assessment report, the provider of TikTok states that it "*manually reviews all content that reaches a certain level of popularity (by monitoring the volume of video views)*" ([49]). Please provide:

a. A non-technical description of the detection mechanism(s) in place for the diverse types of content hosted on TikTok for "rises in popularity" and any other conditions for triggering a further manual review;

125. At the outset TikTok notes, as the Commission is aware: (i) Recital 30 and Article 8 DSA expressly provide that the DSA does not impose, and nothing in it should be construed as imposing, a general monitoring obligation or a general active fact-finding obligation on providers, or a general obligation for providers to take proactive measures in relation to illegal content; and (ii) Article 14(4) requires providers to enforce restrictions in their terms and conditions in a "*diligent, objective and proportionate manner*" with due regard to, among other things, users' fundamental rights such as freedom of expression. TikTok understands that this question does not suggest otherwise.

126. TikTok understands that, in this ***question 9***, the Commission is asking specifically about TikTok's automated mechanisms for detecting that video content uploaded by users to the Platform is "*ris*[ing] *in popularity*" or "*viral content*", and the circumstances in which such content may then be sent for further, manual review.[50]

127. The basis on which TikTok's automated system detects whether a video is "*ris*[ing] *in popularity*" or "*viral content*" is that video's 'video view' ("**vv**") count:

    a. The vv count increases each time a video is viewed (in whole or in part). Whenever a video's vv count reaches one of TikTok's pre-determined thresholds (which are consistent across all markets) – e.g., 20vv, 50vv, 70vv, 100vv, 150vv, 200vv, 250vv, etc. – it is then assessed by a separate Priority Ranking Score model ("**PRS Model**") (see below). That separate assessment determines whether the video will be subject to a further manual review.

    b. Additionally, when a video's vv count reaches TikTok's separate 'high' vv threshold (e.g., 100,000vv – although this varies across markets, depending on what is considered a 'high' vv for that market), the video is also assessed by the PRS Model. In such cases, the Priority Ranking Score calculated by the PRS Model will always be high enough that the video will be sent to human moderation.

128. As explained further in response to ***question 9(d)(ii)*** below, the PRS Model takes into account various factors (including vv as well as, e.g., number of likes, comments and user reports on the video) to assess the video's: (i) risk degree (i.e., the likelihood that the video contains content that violates the Community Guidelines); and (ii) predicted vv count growth (in particular, the extent to which the vv count is predicted to grow in the following seven days). If the video's Priority Ranking Score reaches a certain threshold, the video is then sent for further review by human content moderators.

---

[47] See: https://vsquare.org/far-right-tiktok-lies-conspiracies/ and https://www.isdglobal.org/digital_dispatches/naztok-an-organized-neo-nazi-tiktok-network-is-getting-millionsof-views/, accessed on 26 September 2024.

[48] See: Risk Assessment 2: Integrity & Misinformation. Module 1: Risks to Election & Civic Integrity, registered under Ares(2023)822155, pp. 21 and 38.

[49] See: "TikTok DSA risk assessment report 2024" submitted by the provider of TikTok to the European Commission on 29 August 2024, registered under Ares(2024)6149012, p.15.

[50] See also TikTok's response to RFI Ref.Ares(2023)7112552 (the "Israel-Hamas RFI"), pp. 24-25.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013719

129. For completeness, please note that TikTok uses a similar Priority Ranking Score system in the context of LIVE. Due to the nature of LIVE, which is streaming content, TikTok does not use vv thresholds as a separate trigger for calculating the Priority Ranking Score or for human moderation. Rather, every piece of LIVE content is reviewed to assess a Priority Ranking Score. The Priority Ranking Score for LIVE still considers LIVE room view counts, such that a high LIVE room view count will result in a higher Priority Ranking Score.

b. Clarification on the definition of *"virality"*, detailing whether *"a certain level of popularity"* refers to specific threshold(s) in place, whether popularity varies by type of content, whether popularity is only considered in terms of number of views, or to which extent diffusion patterns are taken into account, including any other factors that may affect the impact that viral content may produce;

130. See TikTok's response to ***question 9(a)*** above with respect to the definition of "*virality*" and "*a certain level of popularity*". Diffusion patterns (i.e., models of the spread of information from person to person on the Platform) are not taken into account in these definitions.

c. A description of the possible actions and outcomes from further manual reviews and, if applicable, a description of performance metrics or internal controls used to assess their effectiveness as risk mitigation measures;

131. TikTok understands that, in this ***question 9(c)***, the Commission is asking specifically about the further manual (or 'human') reviews that are the subject of TikTok's response to ***question 9(a)*** above.

132. Following the human review of a "viral" video as described in response to ***question 9(a)*** above, possible enforcement outcomes include that the video may be:

    a. left as-is and fully available on the Platform;

    b. made FYF-ineligible (see further TikTok's response to ***question 10(a)*** below); or

    c. removed from the Platform.

133. Additionally, action may be taken against the user's account (e.g., they may be banned – see further TikTok's response to ***question 11(c)*** below).

134. Human reviews are conducted as part of TikTok's human content moderation, which in turn is part of TikTok's overall content moderation (including automated content moderation) measures. With respect to that content moderation, relevant metrics are included, e.g., in TikTok's Community Guidelines Enforcement Report[51].

135. Further, TikTok undertakes quality assurance measures to evaluate the extent to which human moderators are: (i) accurate; (ii) apply inappropriate enforcement outcomes/actions; and/or (iii) do not apply appropriate enforcement outcomes/actions. TikTok performs these quality assurance evaluations by including sample cases in human moderators' review queues and comparing the results. These assessments are monitored on a daily, weekly and monthly basis.

[51] See https://www.tiktok.com/transparency/en-gb/community-guidelines-enforcement-2024-9.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013720

d. If automated tools are used to detect *"viral content"* to be manually reviewed, please include:

i. The respective technical system cards or reports, model cards or reports, datasheets or relevant internal documentation for such models, particularly for risk mitigation from the specific recommender systems;

136. As explained in TikTok's response to ***question 9(a)*** above, TikTok determines whether a video is "*viral*" (such that the video may be subject to a manual review) based on whether that video's vv count has reached certain pre-determined thresholds, as set out in response to ***question 9(a)*** above.

137. Videos reaching these thresholds are then assessed by TikTok's PRS Model to determine whether it is sent for review by human content moderators.

138. The documents that TikTok has been able to identify as relevant to this ***question 9(d)(i)***, regarding the PRS Model, are provided as Annexes 070 and 071.

ii. A non-technical description of the training of such models, including a description of the training data used, how such data is selected and labelled, and which performance metric or loss function the models are optimizing for.

139. As explained in TikTok's response to ***question 9(a)*** above, the PRS Model scores content based on (i) its risk score, and (ii) vv growth prediction.

140. The current version of the PRS Model is described in "*Priority Fusion Formula: A Paradigm Shift in Tiktok Video PSF Ranking Mechanism*", provided in response to ***question 9(d)(i).*** This version of the PRS Model is an extension of the model described in "*PSF 2.0: Priority Ranking Model for General Recall Moderation in VMS 3.0"*, also provided in response to ***question 9(d)(i)***.

Training data

141. The PRS Model is trained using the following data:

a. information about the content (e.g., like count, share count, like rate, share rate, report count, and report rate);

b. prediction scores on the probability that the content violates the Community Guidelines, using TikTok's content moderation models;

c. metrics related to the content's vv growth speed and predicted growth (e.g., total video views at that time the Priority Ranking Score mechanism is triggered, the vv growth speed over the last hour or day, time between different vv thresholds); and

d. account-level metrics (e.g., account risk statistics and statistics about the user, including number of comments, number of videos by the user previously removed from the Platform).

Training labels

142. To train the PRS Model, TikTok labels content based on the combined results of the following:

a. 'violative risk score' (corresponding to the probability that the content violates any of the Community Guidelines ("**violative**"));

b. 'unapproved risk score' (corresponding to the probability that the content violates any of the Community Guidelines or is FYF-ineligible (together, "**unapproved**")); and

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013721

c. vv growth predictions.

Loss functions for the PRS Model

143. TikTok uses three loss functions based on binary and multiclass cross entropy loss to measure the accuracy of the PRS Model's predictions, compared to the actual outcomes:

a. *vv based loss function*: TikTok uses a multiclass cross entropy loss function to assess the accuracy of the PRS Model's violative/unapproved vv predictions – i.e., TikTok's risk score assigned to the content (whether violative or unapproved) and its predicted vv growth.

b. *item-based loss function*: TikTok uses a binary cross entropy loss function to assess the accuracy of the PRS Model's predictions about a piece of content's risk score (whether violative or unapproved). To calculate this metric, TikTok compares the content's risk score to its actual moderation outcome where the content was reviewed by human moderators.

c. *weighted loss functions*: TikTok uses a weighted sum combining: (i) the multi-class cross entropy function of the PRS Model's violative/unapproved vv predictions; and (ii) the binary cross entropy function of the PRS Model's prediction of whether a piece of content will be violative or unapproved, to assess the PRS Model's overall performance on vv and risk predictions (whereas the other two loss functions only show such performance individually).

10. In its 2024 risk assessment report, TikTok mentions "*FYF eligibility standards to help ensure that any content promoted by its recommendation system is appropriate for a broad audience*" ([52]). However, clarification is required on the effectiveness of mitigating risks for civic discourse and electoral processes, and on whether eligibility standards also apply to other platform surfaces, such as 'You May Like', 'LIVE', suggested search, suggested accounts, displayed ads, banners, etc. In this context, please provide the following information:

a. A detailed description of all the tests implemented to assess the effectiveness of the "FYF eligibility" standards as a mitigation measure against the risk to civic discourse, media diversity and pluralism, and electoral processes.

144. As explained in TikTok's Year 2 RAR[53], the main feed on the Platform is FYF. FYF is the interface that TikTok's users first see when they open TikTok, and TikTok estimates that its users spend approximately 96% of their time on TikTok scrolling through FYF. Accordingly, the vast majority of content that users see on TikTok is recommended to and can be viewed by them through FYF. TikTok understands that, in this ***question 10(a)***, the Commission is asking specifically about tests implemented by TikTok to assess the effectiveness of the FYF Eligibility Standards (see https://www.tiktok.com/community-guidelines/en/fyf-standards) as a mitigation measure under Article 35 DSA against the systemic risk of actual or foreseeable negative effects on civic discourse and electoral processes and on freedom and pluralism of the media.

145. Under the Civic & Election Integrity section of TikTok's Integrity & Authenticity policy (at https://www.tiktok.com/community-guidelines/en/integrity-authenticity#2), content which is ineligible for recommendation in FYF ("**FYF-ineligible**") – but which TikTok has not (or not yet) assessed to be violative – includes:

a. Unverified Election Claims Policy: Unverified claims about an election, such as a premature claim that all ballots have been counted or tallied (for further information about

[52] See: "*TikTok DSA risk assessment report 2024*" submitted by the provider of TikTok to the European Commission on 29 August 2024, registered under Ares(2024)6149012, p.87.

[53] e.g., p. 13.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

how TikTok fact-checks unverified claims and may make them FYF-ineligible, see the response to ***question 11(b)*** below); and

b. Misrepresented Civic Sources Policy: Statements that significantly misrepresent authoritative civic information, such as a false claim about the text of a parliamentary bill.

146. As noted in the Elections & Civic Integrity Module Questionnaire, taking account of the risks identified in the Year 1 Risk Assessment and evolving risks, TikTok introduced its new Unverified Election Claims and the Misrepresented Civic Sources Policies.

147. Post-launch, TikTok carried out additional testing in January 2024 with a view to assessing the effectiveness of the new policy compared to the pre-existing policy. Results from this testing demonstrated effectiveness of the new policy in slowing the 'rate of video view gain' by 63%.

148. As noted in the Elections & Civic Integrity Module Questionnaire, the introduction of the Misrepresented Civic Sources Policy was completed in early June.

b. All internal documents outlining the results of these tests, including their intended objective, goal, or purpose, as well as any performance metrics or internal controls used to evaluate, measure, or compare the test results. Include any thresholds used to define the results of these tests as effective or successful risk mitigation measures;

149. Copies of the requested documents are provided as Annexes 072 - 073.

150. Details on TikTok's threshold for testing the effectiveness of the risk mitigation measures are provided in TikTok's response to ***question 10(a)***.

151. The FYF Eligibility Standards are one mitigation measure among a wide range of complementary mitigation measures that TikTok implements across the Platform, as explained in this Response and TikTok's Risk Assessment documents.

c. A detailed description of any steps that have been taken to improve the results for future testing.

152. No specific steps have been taken to improve the results referred to in response to ***question 10(a)*** above.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013723

11. The Guidelines for providers of VLOPs and VLOSEs on the mitigation of systemic risks for electoral processes ([54]), formally adopted on 26 April 2024, reference specific measures to mitigate risks posed by recommender systems in the context of electoral processes. In this regard, please reply to the questions below. In case you have already addressed the question above, please reference the specific answer it refers to.

a. How has the provider of TikTok ensured that the specific recommender systems are designed and adjusted in a way that gives users meaningful choices and controls over their feeds, with due regard to media diversity and pluralism? If the provider of TikTok does not employ such mitigation measures, describe how it ensures media diversity and pluralism in its recommender systems and how it mitigates the risk of echo chambers caused by its recommender systems. In the latter case, please provide a timeline for the implementation of the measure.

153. As a preliminary point, TikTok notes that the introduction to this ***question 11*** refers to paragraph 27(d) of Communication C/2024/2537 (the "**Elections Guidelines**"), which contains non-binding recommendations for "*possible measures*" that VLOPs should "*consider*" to mitigate risks in relation to electoral processes within the EU. Without prejudice to this, TikTok provides the following response to this question.

154. TikTok understands that the Commission's question refers in particular to paragraph 27(d)(i) of the Elections Guidelines. As explained below, TikTok has ensured in several ways that its recommender systems are designed and adjusted in a way that empowers users with meaningful choices and controls over their feeds:

a. The goal of TikTok's recommender systems is to help users find content they are likely to be interested in. TikTok recommends content based on several factors to predict how relevant and interesting content might be to a user. The three main factors are user interactions, content information, and user information.

b. TikTok's recommender systems suggest content based on users' preferences as expressed through interactions on TikTok, such as following an account or liking a post. Additionally, when users first sign up for TikTok, TikTok may invite them to select categories of interest, which help form the users' recommendations.

c. In addition to selecting categories of interest when users sign up for TikTok, they can use a number of features to exercise choice over their feeds and the information recommended to them, including for example:

i. *Like*: If a user enjoys specific content they can press like, which helps the systems suggest similar content in the future.

ii. *Not interested*: If a user does not enjoy a specific piece of content, they can let TikTok know by sharing feedback that they are not interested and TikTok will show them less content like that.

iii. *Filter keywords*: Users can use filters to remove content containing specific words and their variations, and hashtags from their FYF and Following feed.

iv. *Refresh*: Users can update the content TikTok recommends on their FYF and interact with popular content, such as liking and commenting, to help reshape their For You feed recommendations.

155. Users can also opt out of the personalised experience for recommender systems based on profiling.

[54] See "*Commission Guidelines for providers of Very Large Online Platforms and Very Large Online Search Engines on the mitigation of systemic risks for electoral processes pursuant to Article 35(3) of Regulation (EU) 2022/2065*", OJ C/2024/3014, 26.04.2024, pp. 1-19.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

156. TikTok has explained in response to ***question 7(a)*** above the measures it has implemented in respect of diversification of content, and in response to ***question 5(c)*** above about the measures it takes to detect, investigate, and disrupt attempts to inauthentically amplify certain viewpoints on the Platform. Further, under TikTok's state-affiliated media ("**SAM**") policy, TikTok applies labels on certain accounts to provide additional context, or access to authoritative information, about civic and elections-related topics. The SAM label appears on content belonging to media organisations whose editorial decision-making or output is subject to control or influence by a government. TikTok identifies where this is the case by working with independent authoritative experts, taking into account a range of information such as the entity's ownership, funding sources and editorial guidelines. By applying the SAM label, TikTok seeks to encourage users to consider, and to ensure users have accurate, transparent, and actionable context about, the source of content they are viewing when they engage with content from media accounts that may present the viewpoint of a government.

b. Which measures has the provider of TikTok established to reduce the prominence of disinformation in the context of elections based on clear and transparent methods? Which factors are used to determine whether such content classifies as disinformation in the context of elections, and how is such assessment made and by whom?

157. TikTok understands that the Commission's question refers in particular to paragraph 27(d)(ii) of the Elections Guidelines. TikTok has established a number of measures to reduce the prominence of Election Misinformation based on clear and transparent methods.

158. **TikTok's policies:** TikTok's Integrity & Authenticity Community Guidelines, which are publicly available and written in plain and intelligible language, include its policies and rules with respect to Misinformation, Civic & Election Integrity, Edited Media and AIGC (the "**AIGC policy**"), and Fake Engagement. TikTok enforces its Community Guidelines and regularly updates them.

a. For example, ahead of the EU Elections, TikTok updated its Synthetic and Manipulated Media policy to become what is now the AIGC policy. Through those updates, TikTok adopted more commonly used and easily understood language in the context of AIGC, and clarified the existing prohibitions on AIGC showing fake authoritative sources or crisis events, or falsely showing public figures in certain contexts. See further TikTok's response to the GenAI RFI.

b. As further explained in the Elections & Civic Integrity Module Questionnaire, and in TikTok's Year 2 RAR, TikTok has also expanded its policy coverage and updated its Community Guidelines in a number of areas. As part of this process, TikTok launched:

i. The Election Misconduct policy, which targets content that promotes or provides instruction on illegal participation and electoral interference, or calls for the disruption of a legitimate outcome of an election outside of the legal system, such as through a coup;

ii. The Distribution of Hacked Materials Policy, which was designed to meet gaps identified through TikTok's Year 1 risk assessment process, and prohibits the sharing of hacked materials that may lead to disturbances in civic processes, escalation of violence in conflict zones, polarisation in society, or reduction of trust in the information ecosystem; and

iii. The Misrepresented Civic Sources policy, which instructs specialised moderator queues to restrict false content that significantly misinforms users about key sources of civic information which may harm the information ecosystem, undermine public participation in civic processes, and reduce awareness of political issues.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013725

c. TikTok's Safety and Content Advisory Council, which includes independent experts in safety-related areas such as freedom of expression, provides guidance to help TikTok develop forward-looking policies that address existing challenges and anticipate those of the future.

159. In addition to its publicly available Community Guidelines, TikTok provides other transparent and publicly available resources through its Transparency Centre and Help Centre. These include explanations about TikTok's approach to content moderation[55], protecting election integrity[56], and combating harmful misinformation through TikTok's Fact-Checking Program and other means[57]. TikTok also produces regular transparency reports which are available through the Transparency Centre (which meet its relevant transparency reporting obligations under the DSA). Those reports include Community Guidelines Enforcement Reports[58] and reports on TikTok's work to detect and disrupt CIOs[59].

160. TikTok's approach to enforcement of its Community Guidelines to reduce the prominence of Election Misinformation, including detecting and acting against CIOs, is further explained in its Year 1 and 2 RARs and in response to ***question 3 - 10*** above. Notably, in the first half of 2024, 97% of videos violating TikTok's misinformation policies were removed proactively, before any user report.

161. **TikTok's fact-checking:** As explained above, TikTok uses human and automated means to detect, identify and take appropriate content moderation action against Election Misinformation. As explained in response to ***question 6(a)*** above, content identified as potential Election Misinformation is reviewed by specially-trained human moderators, who review and assess the content of the video, taking into account TikTok's Integrity & Authenticity Community Guidelines. If the human moderators are unable to determine the veracity of a potential Election Misinformation claim by reference to TikTok's database of previously fact-checked claims, they can escalate it to one of TikTok's independent fact-checking partners[60], who conduct a further assessment and advise TikTok how to classify the information. TikTok then takes the appropriate content moderation action (e.g., removing confirmed Election Misinformation, and making unverifiable claims FYF-ineligible to restrict their prominence in users' FYF).

162. TikTok also prompts users to reconsider sharing certain content that is unverified (see TikTok's response to ***question 6(a)*** above) and potentially harmful Election Misinformation and displays reminders to users concerning Election Misinformation, such as regarding the ability to use TikTok's reporting mechanisms to report potential Election Misinformation for consideration by TikTok's content moderation teams. TikTok's TnS teams action reports of violative content made by users, non-users and by third parties who form part of TikTok's Community Partner Channel (similar to trusted flaggers under the DSA). To date, more than 400 organisations who specialise in a range of safety issues use the Community Partner Channel.

163. Additionally, expert teams at TikTok investigate and take action against CIOs which may seek to spread Election Misinformation. These expert teams analyse behaviour and account linkages, monitor cross-Platform threats, manually investigate reports and, where CIOs are identified, take enforcement action to disrupt them. In addition, suspicious accounts and behaviours are automatically removed, and inauthentic engagement signals are ignored in estimating account or video popularity. See further TikTok's responses to ***question 5(c)*** and ***question 9*** above.

164. **COPD and the EU Elections:** Since 2020, TikTok has actively participated in the Code of Practice on Disinformation's ("**COPD**") Taskforce and working groups, including co-chairing the Elections subgroup and Transparency subgroup, to stay at the forefront of combating Election Misinformation, including through collaboration with the Commission and signatories on trends and mitigation strategies. TikTok's holistic approach includes preventing the spread of Election Misinformation,

[55] See https://www.tiktok.com/transparency/en/content-moderation.
[56] See https://www.tiktok.com/transparency/en/protecting-elections.
[57] See https://www.tiktok.com/transparency/en/combating-misinformation.
[58] See https://www.tiktok.com/transparency/en-us/community-guidelines-enforcement-2024-9.
[59] See https://www.tiktok.com/transparency/en-us/covert-influence-operations.
[60] See further https://www.tiktok.com/transparency/en-us/combating-misinformation/#GFCP.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013726

elevating reliable election-related information, and investing in media literacy to help build resilience across TikTok's community.

165. TikTok regularly publishes its publicly available COPD Reports, which explain in detail the measures put in place against Election Misinformation, including in the context of elections. TikTok's September 2024 COPD Report explains that, beyond the measures set out above, in relation to the EU Elections, key mitigation measures implemented to reduce the prominence of Election Misinformation included:

a. Mission Control Centre ("**MCC**"): TikTok established a dedicated MCC in its Dublin office, bringing together in-person employees from multiple specialist teams to provide consistent and dedicated coverage of potential elections-related issues in the run-up to, and during, the EU Elections. In the four weeks leading up to and including the EU Elections, TikTok removed over 43,000 pieces of content for violating TikTok's misinformation policies, proactively removing over 96% of violative misinformation content before it was reported, and over 80% before it received a single view by users; and

b. Rapid Response System ("**RRS**"): TikTok actively participated in the setup of the COPD RRS, launched in April 2024 to streamline the exchange of information among civil society organisations, fact-checkers, and online platforms. The RRS was a dedicated framework for cooperation among signatories during the EU Elections, which allowed non-platform signatories to flag time-sensitive content, accounts or trends that may present threats to the integrity of the electoral process. TikTok received five notifications from civil society organisations through the RRS before the EU Elections, which were rapidly actioned and resulted in banning or geo-blocking of accounts and content removals for violation of Community Guidelines.[61]

c. Has the provider of TikTok adopted measures to reduce the prominence of disinformation in the context of elections, specifically coming from accounts that have repeatedly been found to spread disinformation? If yes, which ones?

166. TikTok understands that the Commission's question refers in particular to paragraph 27(d)(ii) of the Elections Guidelines. Further to TikTok's response to ***question 11(b)*** above:

167. As described in TikTok's Help Centre articles about Accounts & Features[62] and Content Violations and Bans,[63] violating TikTok's Community Guidelines can lead to enforcement actions on accounts. TikTok will ban accounts or account holders that engage in:

a. A single severe content violation;

b. Repeated content violations; or

c. Circumventing enforcement of, or operating accounts dedicated to activity that violates TikTok's Community Guidelines.

168. TikTok adopts a tiered approach to enforcement of these rules based on "strikes", as follows:

[61] Further detail can be found in TikTok's Newsroom article at https://newsroom.tiktok.com/en-eu/tiktoks-continued-commitment-to-combating-disinformation-under-the-eu-code-of-practice, and in TikTok's September 2024 COPD Report at https://disinfocode.eu/signatory-report/tiktok-september-2024-report/.

[62] See https://www.tiktok.com/community-guidelines/en/accounts-features.

[63] See https://support.tiktok.com/en/safety-hc/account-and-user-safety/content-violations-and-bans.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

a. The first time a user's violative content is removed: (i) if the violation is severe, the user may receive an immediate strike or even immediate ban; (ii) otherwise, the user will receive a "warning strike".

b. For each future violation, the account receives a "strike". TikTok's content moderation system counts the number of strikes, by Community Guidelines policy area (for example, Integrity & Authenticity, which covers Election Misinformation) and by feature (for example, comments) and the account will receive a strike based on the severity of the policy violation. Repeat violators tend to follow a pattern – TikTok's analysis has found that almost 90% violate using the same feature consistently, and over 75% violate the same policy category repeatedly. TikTok notifies users of any consequences of violations, including if they are on the verge of being permanently banned. Users can also access a record of their violations. Strikes expire from an account's record after 90 days.

c. In some cases, for violations involving certain features such as LIVE or direct messages, TikTok may temporarily restrict the user's access to the feature while the content is under review, to ensure that the user doesn't immediately re-engage in violative behaviour through that Platform feature.

d. When the user's strikes reach a certain threshold (depending, e.g., on the particular Community Guidelines policy that has been repeatedly violated), TikTok may permanently ban the account. Owners of accounts that have been banned or restricted may try to create another account on TikTok to avoid those penalties. In such cases, TikTok may restrict or permanently ban the new account as well.

169. An exception applies for GPPPAs (and other such "**public interest accounts**", including news entities), as explained in response to ***question 5(e)*** above.

a. Although TikTok uses the same content moderation practices for GPPPAs as for "regular" TikTok accounts (e.g., removing any violative content and permanently removing the account for any single severe content violation, such as showing real-world violence or torture), TikTok recognises that GPPPAs play a particular role in civic processes and civil society. Therefore, in accordance with TikTok's commitments and obligations in respect of users' fundamental rights and free expression, different enforcement rules apply to public interest accounts that repeatedly post violative content. If such accounts reach the strike limit, instead of being banned, they will be temporarily ineligible to appear in FYF and Following feeds for 90 days.

b. Where public interest may present a particularly high risk to public safety, e.g., during election periods and other high-risk social and political events, TikTok may impose additional or different restrictions. During such periods, if a public interest account posts content that promotes violence, hate, or misinformation that could undermine a civic process or contribute to real-world harm (including Election Misinformation), then TikTok may restrict that account from posting content for a period of 7 to 30 days, depending on the severity of the violation and surrounding risk.

c. TikTok may also extend the restriction period if it considers that the actions of the account owner indicate they are likely to continue violations, and public safety is still at high risk. In making that decision, TikTok may also take into account the public interest account owner's behaviour outside of TikTok.

170. In addition to the above penalties for repeatedly posting violative content, TikTok may make an account and its content FYF-ineligible, and harder to find in Search, where TikTok identifies that the account is repeatedly posting FYF-ineligible content (see further TikTok's response to ***question 10(a)*** above).

171. Finally, further to its response to ***question 5(c)*** and ***question 11(b)*** above, TikTok notes that its CIO teams continuously monitor Election Misinformation campaigns, inauthentic behaviour and CIOs.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013728

The teams identify and take action against accounts that are committed to and/or repeatedly attempting to manipulate or corrupt public debate and/or spread Election Misinformation.

d. Has TikTok regularly assessed the performance and impact of its specific recommender systems and addressed any emerging risks or issues related to electoral processes? If yes, how?

172. TikTok understands that the Commission's question refers in particular to paragraph 27(d)(iv) of the Elections Guidelines.

173. TikTok has regularly assessed the performance and impact of its recommender systems and has identified, and implemented mitigation measures to address, emerging risks related to electoral processes – including by updating and refining policies, practices, and algorithms. For example, in its Year 2 risk assessment, TikTok identified a number of new risk areas that had emerged following its Year 1 risk assessment, including risks related to potentially harmful content being recommended by its recommender systems in the context of elections and civic processes. Those risks, and mitigation measures that TikTok has implemented in respect of them, are explained in the Elections & Civic Integrity section of the Year 2 RAR, and in the Elections & Civic Integrity Module Questionnaire. They include, e.g.:

a. Risks of unverified claims that may harm users' trust in elections/civic processes being shared and amplified on the Platform. TikTok has addressed this risk by, e.g., updating its Community Guidelines following consultations with over 40 global experts in democracy, tech policy and freedom of speech, adding new policies and enforcement regarding unverified elections claims, election misconduct, distribution of hacked materials and misrepresented civic sources. TikTok also worked with behavioural psychologists Irrational Labs, which led to the development of warning and labelling features, including to help users consider before potentially amplifying unverified content; and

b. Risks from evolving SAM activities, such as using influencers to amplify state-affiliated narratives and/or using individual media employees to propagate controlled messaging. In respect of such risks, TikTok, among other things, implements its SAM policy (discussed in response to ***question 11(a)*** above). TikTok applies a prominent label to all content and accounts from SAM. The user is shown a screen pop-up providing information about what the SAM label means, inviting the user to "learn more", and redirecting them to an in-app page. The measure brings transparency, raises users' awareness, and encourages users to consider the reliability of the source of the content they are being recommended. To inform the development and implementation of its SAM policy, TikTok consults with media experts, political scientists, academics, and representatives from international organisations and civil society across the world, including numerous academics in Europe.

c. Additional potentially harmful types of Election Misinformation which may be recommended on the Platform.

This was identified through a combination of expert consultations, media monitoring, internal expertise, preparations for the 2024 elections, and engagement with relevant communities and events. To help counter Election Misinformation, as described in TikTok's responses to ***questions 11(b) and 11(c)*** above, TikTok regularly consults with third party experts and researchers in relation to the development and update of policies and features which are designed to reduce the spread of disinformation. For example, TikTok engaged with experts globally on TikTok's Election Misinformation policies, which help inform updates of TikTok's Integrity & Authenticity policies. Additionally, TikTok's Global Security Office ("**GSO**") regularly reviews recommender system code and configurations to check for potential bias in respect of GPPPAs.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013729

e. Which measures has the provider of TikTok established to provide transparency around the design and functioning of its specific recommender systems, in relation to the data and information used in designing systems that foster media pluralism and diversity of content, to facilitate third party scrutiny and research?

174. TikTok refers to and repeats the comments at the beginning of its response to ***question 11(a)*** above regarding the basis of ***questions 11(a)-(f)***. TikTok understands that the Commission's question refers in particular to paragraph 27(d)(v) of the Elections Guidelines and is not asked with reference to TikTok's compliance with Art. 40(12) DSA.

175. TikTok has previously provided the Commission with detailed information concerning its Research Tools and Virtual Compute Environment. See further: (i) TikTok's responses to RFI Ref.Ares(2024)398592 (the "**Public Data Access RFI**"); (ii) TikTok's responses to questions 68-73 of the Youth Investigation RFI; and (iii) TikTok's resources for researchers at https://developers.tiktok.com/products/research-api/.

176. TikTok's Research Tools supports research in areas such as misinformation, disinformation, violent extremism, social trends, and community building. The Research Tools allows researchers with the necessary expertise who conduct research on a not-for-profit basis to access certain data, including the following information about public account owners.

    a. Videos: Videos that can be watched by "Everyone" and, for such videos, the total number of likes, total number of comments, total number of favourites, voice-to-text, subtitles, whether it can appear in TikTok's STEM, associated metadata (e.g. the time of creation), playlist information, as well as associated user content such as hashtags.

    b. Comments: Comment text, and total number of likes, replies, and the time the comment was posted.

    c. Accounts: Bios, profile pictures, country of registration (for example, a region code such as IE), the usernames and profile names, URLs shared by users who are verified, and engagement data (e.g. following list and followers list, and shop performance where the account sells on TikTok Shop).

177. TikTok also provides transparent, public information about the design and functioning of its recommender systems in its Help Centre and Transparency Centre, including at:

    a. https://support.tiktok.com/en/using-tiktok/exploring-videos/how-tiktok-recommends-content;

    b. https://www.tiktok.com/transparency/en-us/recommendation-system; and

    c. https://www.tiktok.com/transparency/en-us/system-guides.

f. Has the provider of TikTok engaged with external parties to conduct adversarial testing and red team exercises on these systems to identify potential risks such as risks stemming from biases, susceptibility to manipulation, or amplification of misinformation, disinformation, Foreign Information Manipulation and Interference, or other harmful content. In the case such testing has taken place, what were the results and how were learnings implemented? If yes, how?

178. TikTok understands that the Commission's question refers in particular to paragraph 27(d)(vi) of the Elections Guidelines.

179. As the Commission is aware, ahead of the EU Elections, TikTok's teams participated in events conducted by the Commission. In particular, in April 2024, TikTok took part in a simulation exercise to test TikTok's election readiness. That exercise was attended by civil society organisations and a number of European authorities. TikTok also took part in dedicated EU Elections roundtables in May

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

and June 2024, and a post-election roundtable in July 2024. The roundtables offered valuable opportunities to learn about actions peer industry members were taking to address common risks faced during elections and to learn about the Commission's concerns and priorities.

180. As well as external engagement, an internal exercise was conducted to identify opportunities and gaps at a process and operational level, simulating escalations and scenarios specific to the election. As part of that exercise, particular markets and team members were selected to stress-test several cases taking into account aspects like high volume, reduced capacity and support, as well as the scenario planning document and response plan that TikTok's Policy team built.

Schedule of Annexes

The below Schedule contains a list of documents identified as responsive to the requests in this RFI. TikTok further notes that:

- There are two responsive documents identified as being either privileged or partially privileged. Therefore, for the purposes of protecting such privilege in other jurisdictions and contexts, and on the advice of external counsel, TikTok objects to the disclosure of such privileged information. We are available at your convenience to discuss.
- The documents listed in this Schedule necessarily reflect matters at the time the documents were created (or last updated, where applicable). As such, some of the information they contain may not fully reflect the current position. Accordingly, please refer to TikTok's responses to the questions of this RFI for the position as at the date of submission of those responses.
- Any documents listed in this Schedule have been redacted to remove the following information: (i) personal information; (ii) information that clearly relates solely to countries outside the EEA and is confidential and commercially sensitive; and/or (iii) legally privileged information.
- TikTok is producing documents under applicable laws but not waiving privilege claims that may be applicable in any jurisdictions. In the event that any waiver of privilege does nonetheless occur, such waiver is limited to the disclosure of the relevant document(s) to the Commission only and only for the purposes of our response to this RFI.

Question No.	Annex No.	Document	Privileged	Confidential	Redacted
1(a)	001	Updated October 2024 - Year 2 DSA Risk Assessment report	No	Yes	Yes
1(a)	002	Year 2 SRA: Analysis, assessment and scoring of systemic risk	No	Yes	No
1(a)	003	[MASTER - FINAL] Questionnaire for DSA Risk Assessment - Age Appropriate Content	No	Yes	Yes
1(a)	004	[MASTER - FINAL] Questionnaire for DSA Risk Assessment - Elections & Civic Integrity.pdf	No	Yes	Yes
1(a)	005	[MASTER - FINAL] Questionnaire for DSA Risk Assessment - Gender Based Violence	No	Yes	Yes
1(a)	006	[MASTER - FINAL] Questionnaire for DSA Risk Assessment - Online Engagement (Wellbeing)	No	Yes	Yes
1(a)	007	[MASTER - FINAL] Questionnaire for DSA Risk Assessment - Terrorist Content	No	Yes	Yes
1(a)	008	[MASTER - FINAL] Questionnaire for DSA Risk Assessment - TNS/GBS IP Ops	No	Yes	Yes
1(a)	009	[MASTER - FINAL] Questionnaire for DSA Risk Assessment Harmful Misinformation.pdf	No	Yes	Yes
1(a)	010	[MASTER - FINAL] Questionnaire for DSA Risk Assessment Medical Misinformation.pdf	No	Yes	Yes

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013732

Question No.	Annex No.	Document	Privileged	Confidential	Redacted
1(a)	011	[MASTER - FINAL] Questionnaire for DSA Risk Assessment: CSAM and CSEA	No	Yes	Yes
[redacted]	012	[MASTER - FINAL] Questionnaire for DSA Risk Assessment_Fundamental Rights	No	Yes	Yes
1(a)	013	[MASTER- FINAL] Questionnaire for DSA Risk Assessment - Age Assurance	No	Yes	Yes
1(a)	014	[MASTER- FINAL] Questionnaire for DSA Risk Assessment - Hate Speech	No	Yes	Yes
1(a)	015	Age Assurance Risk Assessment Review Group Minutes	No	Yes	Yes
1(a)	016	Age-Appropriate Content Risk Assessment Review Group Minutes	No	Yes	Yes
1(a)	017	CSAM Risk Assessment Review Group Minutes	No	Yes	Yes
1(a)	018	Elections and Civic Integrity Risk Assessment Review Group Minutes	No	Yes	Yes
1(a)	019	Fundamental Rights Risk Assessment Review Group Minutes	No	Yes	Yes
1(a)	020	Gender-based Violence Risk Assessment Review Group Minutes	No	Yes	Yes
1(a)	021	Harmful Misinformation Risk Assessment Review Group Minutes	No	Yes	Yes
[redacted]	022	Hate Speech Risk Assessment Review Group Minutes	No	Yes	Yes
1(a)	023	IP-Infringing Content Risk Assessment Review Group Minutes	No	Yes	Yes
1(a)	024	Medical Misinformation Risk Assessment Review Group Minutes	No	Yes	Yes
1(a)	025	Online Engagement Risk Assessment Review Group Minutes	No	Yes	Yes
1(a)	026	Terrorist Content Risk Assessment Review Group Minutes	No	Yes	Yes
1(b)	027	[HHB] Toxic Masculinity - Filter Bubble Test	No	Yes	Yes
1(b)	028	[Launch Review][Global] Reduce CGVR in All User's FYP Using New Harassment & Hateful Behavior Model(Multi-issue v2 Version) via Post-Recall Demotion	No	Yes	Yes

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Question No.	Annex No.	Document	Privileged	Confidential	Redacted
1(b)	029	[P&C] [Hate] [PRD] Toxic Gender Narratives - Filter bubble Detection & Dispersion	Partially	Yes	Yes
1(b)	030	[P&C] Internal TT Human Rights Policy	No	Yes	Yes
1(b)	031	[T&S] [HHB] Dispersion of Toxic Narratives on Protected Groups	No	Yes	Yes
1(b)	032	150724 - Minutes - ESAC interview with Alex Holmes	No	Yes	Yes
1(b)	033	A1 - Our continued actions to protect the TikTok community during the Israel-Hamas war _ TikTok Newsroom.pdf	No	No	No
1(b)	N/A	AI Fairness Guidelines 2.0 - P&C	Yes	Yes	N/A
1(b)	034	AI Fairness Review Process 1-Pager	No	Yes	Yes
1(b)	035	DCU-Toxicity-Full-Report.pdf	No	No	No
1(b)	036	Grey-Test Internal System Operation Handbook 2024	No	Yes	Yes
1(b)	037	RWH x FQ Hate - WIP	No	Yes	Yes
1(b)	038	Specialist Teams	No	Yes	No
1(b)	039	T&S Policy TikTok Community Charter	No	Yes	Yes
1(b)	040	Understanding Toxicity for Feed!	No	Yes	Yes
1(c)	041	22 Aug 2024 - TT Ireland Board Meeting.pptx	No	Yes	Yes
4(a)	042	[T&S] EU Election 2024 - BG Election Center (H5) PRD	No	Yes	Yes
4(a)	043	[T&S] EU Election 2024 - CZ Election Center (H5) PRD	No	Yes	Yes
4(a)	044	[T&S] EU Election 2024 - DE/AT Election Center (H5) PRD	No	Yes	Yes
4(a)	045	[T&S] EU Election 2024 - DK Election Center (H5) PRD	No	Yes	Yes
4(a)	046	[T&S] EU Election 2024 - EE/LV/LT Election Center (H5) PRD	No	Yes	Yes
4(a)	047	[T&S] EU Election 2024 - Election Center (H5) PRD	No	Yes	Yes

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013734

Question No.	Annex No.	Document	Privileged	Confidential	Redacted
4(a)	048	[T&S] EU Election 2024 - ES Election Center (H5) PRD	No	Yes	Yes
4(a)	049	[T&S] EU Election 2024 - FI Election Center (H5) PRD	No	Yes	Yes
4(a)	050	[T&S] EU Election 2024 - FR Election Center (H5) PRD	No	Yes	Yes
4(a)	051	[T&S] EU Election 2024 - GR Election Center (H5) PRD	No	Yes	Yes
4(a)	052	[T&S] EU Election 2024 - HR Election Center (H5) PRD	No	Yes	Yes
4(a)	053	[T&S] EU Election 2024 - HU Election Center (H5) PRD	No	Yes	Yes
4(a)	054	[T&S] EU Election 2024 - IT Election Center (H5) PRD	No	Yes	Yes
4(a)	055	[T&S] EU Election 2024 - LU Election Center (H5) PRD	No	Yes	Yes
4(a)	056	[T&S] EU Election 2024 - NL Election Center (H5) PRD	No	Yes	Yes
4(a)	057	[T&S] EU Election 2024 - PL Election Center (H5) PRD	No	Yes	Yes
4(a)	058	[T&S] EU Election 2024 - PT Election Center (H5) PRD	No	Yes	Yes
4(a)	059	[T&S] EU Election 2024 - RO Election Center (H5) PRD	No	Yes	Yes
4(a)	060	[T&S] EU Election 2024 - SE Election Center (H5) PRD	No	Yes	Yes
4(a)	061	[T&S] EU Election 2024 - SI Election Center (H5) PRD	No	Yes	Yes
4(a)	062	[T&S] EU Election 2024 - SK Election Center (H5)	No	Yes	Yes
4(a)	063	[T&S] EU Election 2024 - Video Notice Tag PRD	No	Yes	Yes
4(a)	064	[T&S] EU Election Search Guide PRD	No	Yes	Yes
4(a); 4(b)	065	[T&S] EU Elections Centre H5 Feedback	No	Yes	Yes
4(a)	066	[T&S] EU/Local Election 2024 - IE Election Center (H5) PRD	No	Yes	Yes

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Question No.	Annex No.	Document	Privileged	Confidential	Redacted
4(a)	067	[T&S] EU/National Election 2024 - BE Election Center (H5) PRD	No	Yes	Yes
4(a)	068	[T&S] MT/CY Election 2024 - Election Center (H5) PRD	No	Yes	Yes
4(a); 4(b)	069	[TnS] LIVE Room Banner for EU Election 2024	No	Yes	Yes
9(d)(i)	070	Priority Fusion Formula: A Paradigm Shift in Tiktok Video PSF Ranking Mechanism	No	Yes	Yes
9(d)(i)	071	PSF 2.0: Priority Ranking Model for General Recall Moderation in VMS 3.0	No	Yes	Yes
10(b)	072	Product Policy Proposal - Unverified Election Claims	No	Yes	Yes
10(b)	073	Y2 DSA SRA Data Annex [FINAL]	No	Yes	No

Exhibit 296

EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR COMMUNICATIONS NETWORKS, CONTENT AND TECHNOLOGY

Platforms Policy and Enforcement
Digital Services

Brussels, 2 October 2024

SENSITIVE (*)

Point of contact, via email: [redacted]@TikTok.com

Redacted via email: Redacted tiktok.com

Subject: Request for information pursuant to Article 67(2) and Article 34(3) of Regulation (EU) 2022/2065 regarding TikTok's compliance with Articles 34 and 35 of Regulation (EU) 2022/2065

Dear Redacted

By decision of 25 April 2023, TikTok was designated as a very large online platform ('VLOP') pursuant to Article 33(4) of Regulation (EU) 2022/2065 (the Digital Services Act) ([1]). On the following day, the designation decision was notified to TikTok Technology Limited ("TikTok Ltd.") as the main establishment of the provider of TikTok in the European Union. An acknowledgement of receipt was received by the Commission on 27 April 2023. A copy of the designation decision was sent to ByteDance Limited ("ByteDance"), the company controlling the group of legal entities to which TikTok Ltd. belongs (hereinafter 'the provider of TikTok').

The Commission considers that the provider of TikTok consists of both TikTok Ltd., as the main establishment of the provider within the meaning of Article 56(1) of Regulation (EU) 2022/2065, and ByteDance as the legal entity controlling the group of legal entities to which TikTok Ltd. belongs. Pursuant to Articles 33(6) and 92 of Regulation (EU) 2022/2065, that Regulation started to apply to TikTok as of on 28 August 2023.

Pursuant to Article 67(2) of Regulation (EU) 2022/2065 the Commission may, by simple request, require the provider of a VLOP to provide information on a suspected infringement of that Regulation. In accordance with Article 65(1), the Commission services may make such a request before initiating formal proceedings.

(*) Distribution only on a need-to-know basis. Do not read or carry openly in public places. Must be encrypted in transmission and stored securely, where possible by using encrypted storage. Destroy copies by shredding or secure deletion. Full handling instructions: https://europa.eu/!db43PX

([1]) See: Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance) OJ L 277, 27.10.2022, p. 1–102.

Pursuant to Article 34(3) of Regulation (EU) 2022/2065, providers of VLOPs are required to preserve the supporting documents of the risk assessments for at least three years after the performance of the risk assessments and to communicate them to the Commission, upon request. That provision further provides that the Commission may request such supporting documentation on which the platform relied to draw up its risk assessment report.

For the reasons set out below, the Commission suspects that the provider of TikTok may not be compliant with certain obligations under Regulation (EU) 2022/2065. By the present letter, the Commission therefore requests certain information pursuant to Article 67(2) of that Regulation, as well as the documentation supporting TikTok's risk assessment pursuant to Article 34(3) of that Regulation.

Pursuant to Article 34(1) of Regulation (EU) 2022/2065, providers of VLOPs are obliged to diligently identify, analyse and assess any systemic risks in the Union stemming from the design or functioning of their service and its related systems, including algorithmic systems, or from the use made of their services. Pursuant to Article 34(2) of Regulation (EU) 2022/2065, in their risk assessment providers of VLOPs must take into account some specific factors that can influence any of the systemic risks referred to in Article 34(1), including the design of their recommender systems.

Furthermore, pursuant to Article 35(1) of Regulation (EU) 2022/2065, providers of VLOPs are obliged to put in place reasonable, proportionate and effective mitigation measures tailored to the specific systemic risks identified. Such measures may include: (i) adapting the design, features or functioning of their services, including their online interfaces; (ii) adapting their terms and conditions; (iii) testing and adapting their algorithmic systems, including their recommender systems; (iv) reinforcing the internal processes, resources, testing, documentation, or supervision of any of their activities in particular as regards detection of systemic risks; (v) taking awareness-raising measures and adapting their online interface in order to give recipients of the service more information; (vi) taking targeted measures to protect the rights of the child.

On 29 September 2023, the provider of TikTok submitted its risk assessment report to the Commission. On 29 August 2024, the provider of TikTok submitted the second risk assessment report ([2]). The reports also include information about mitigation measures under Article 35(1) of Regulation (EU) 2022/2065.

Based on these reports, the Commission suspects that the provider of TikTok has not complied with all the obligations under Articles 34 and 35 of Regulation (EU) 2022/2065, in particular with regards to the diligence with which it identified, analysed and assessed the systemic risks related to the design and functioning of the recommender systems within TikTok. Similarly, the Commission suspects, based on the information provided in the report, that the measures adopted to mitigate certain risks stemming from the recommender systems within TikTok may not be reasonable, proportionate and effective.

([2]) See: "TikTok DSA risk assessment report", submitted by the provider of TikTok to the European Commission on 29 September 2023, registered under Ares (2023)6749619; "TikTok DSA risk assessment report 2024" submitted by the provider of TikTok to the European Commission on 29 August 2024, registered under Ares(2024)6149012.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

While TikTok's risk assessment reports provide information on the results related to the identification, analysis and assessment of the risks stemming from the design or functioning of its recommender system in the 'For You Feed', it does not provide any information on the results related to the identification, analysis and assessment of risks stemming from the design or functioning of its recommender systems in 'You May Like' and 'Following' as well as in its search suggestions. As further explained in the questions in Annex I below, the Commission therefore has reasons to suspect that the provider of TikTok has not diligently identified, analysed and assessed whether and how the design of TikTok's recommender systems and any other relevant algorithmic system may influence the risks of actual or foreseeable negative effects of the design or functioning of these recommender systems, especially on civic discourse, electoral processes, and pluralism of the media, as well as how the systemic risks pursuant to Article 34(1) of Regulation (EU) 2022/2065 are influenced by intentional manipulation of its service, including by inauthentic use or automated exploitation of the service.

Furthermore, while TikTok's risk assessment reports provide information on systemic risks pursuant to Article 34 of Regulation (EU) 2022/2065, that might stem from their recommender system 'For You Feed' and on measures that were taken to mitigate these risks pursuant to Article 35 of Regulation (EU) 2022/2065, the Commission has reasons to suspect that these measures might not be effective to mitigate risks in relation to the design or functioning of these recommender systems, especially for civic discourse, electoral processes, and pluralism of the media, and that in its risk assessments the provider of TikTok might not have diligently considered how these are influenced by intentional manipulation of its service, including by inauthentic use or automated exploitation of the service.

Based on Articles 67(2) and 34(3) of Regulation (EU) 2022/2065, the Commission requests the provider of TikTok to reply to the questionnaire attached and to submit the supporting documents requested in Annex I, in order to enable the Commission to further assess whether the provider of TikTok complies with Articles 34 and 35 of Regulation (EU) 2022/2065, no later than **15 November 2024**.

Please submit your reply by sending it to the functional mailbox of the Directorate General for Communications Networks, Content and Technology at CNECT-DSA-Registry@ec.europa.eu.

Unless you specify otherwise, your response shall be treated as non-confidential. If you request that any piece of information supplied in your response should not be published in any public version of Commission documents related to TikTok's compliance with Regulation (EU) 2022/2065 or otherwise disclosed to third parties, please submit both a confidential and a non-confidential version of your reply and provide reasons for any confidentialities claimed.

Any personal data submitted in your reply will be processed in compliance with Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 ([3]).

([3]) See: Regulation (EU) 2018/1725 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC (OJ L 295, 21.11.2018, p. 39).

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014747

Please see here the privacy statement on the application of the regulatory framework, including processing of notifications, monitoring of compliance, supervisory activities, and other activities under the Digital Services Act.

If you consider that some documents included in your reply (or parts of some documents) contain personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership or genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation ([4]) (i.e. special categories of personal data), please provide those documents in a separate submission identified as "SPECIAL CATEGORIES OF PERSONAL DATA", linking each document to the question number to which the document constitutes a reply. Please highlight the special categories of personal data in the document and provide an explanation as to the specific category in case this is not immediately obvious from the information highlighted.

Please note that, pursuant to Article 74(2) point (a) of Regulation (EU) 2022/2065, the Commission may impose on a provider of a VLOP fines not exceeding 1 % of the total annual income or worldwide turnover in the preceding financial year where they intentionally or negligently supply incorrect, incomplete or misleading information in response to a simple request for information pursuant to Article 67.

If you have any further questions, please address them by e-mail to CNECT-DSA-Registry@ec.europa.eu, clearly indicating the subject of this letter and the reference number.

Yours faithfully,

Redacted

([4]) See: Article 10(1) of Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data (EU-DPR), and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC (OJ L 295, 21.11.2018, p.39), which is identical to Article 9(1) of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (GDPR) (OJ L119, 4.5.2016, p.1).

Enclosures:

– Annex I: Questionnaire

– Annex II: Instructions for replying

c.c.: [redacted]@tiktok.com; Redacted @bytedance.com; Redacted bytedance.com

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

ANNEX I

QUESTIONNAIRE

DEADLINE FOR REPLY: 15 November 2024

1. BUSINESS SECRETS AND OTHER CONFIDENTIAL INFORMATION

Does your reply contain any parts that you regard as business secrets or other confidential information that should not be published or otherwise disclosed to third parties?

YES ☐ NO ☐

If yes, please submit both a confidential and a non-confidential version of your reply and provide reasons for any confidentialities claimed.

2. QUESTIONS

Format of the response

See Annex II. Please provide your response in Word or Excel or in machine-readable and searchable PDF.

Specific questions

The questions are grouped by issue as reported in the cover letter. In some questions you are requested to provide a "list of documents". This means that for each answer, the Commission requires you to submit all relevant documents, including documents about the effectiveness of all relevant measures, internal controls, relevant screenshots, presentations to management, internal exchanges/emails about effectiveness of the measures, meeting notes, and other measures considered and discarded, and including any relevant documentation maintained pursuant to Article 34(3) of Regulation (EU) 2022/2065.

In several questions you are requested to provide information on *performance metrics* and *internal controls*:

- By *performance metrics*, the Commission means in particular, but not only, quantitative performance indicators that help determine objectively whether a given policy, action or measure is successful or not, or that allow the determination of progress towards reaching the objectives of the measure.

- By *internal controls*, the Commission means any type of measure, including business process and tests conducted and organisation arrangement, that aims to monitor and ensure the effectiveness and correct functioning of the measures put in place.

Where your organisation is using different terminology, please explain the relevant organisational set-ups that are closest to the requested concepts.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014750

Period of time covered

Unless otherwise specified, the period of time covered by this request for information is the period of time starting as from the date from which Regulation 2022/2065 started to apply to TikTok, i.e. as from 28 August 2023, until the closest possible date for which data/information were available before the submission of your reply. This period is hereinafter referred to as 'the relevant period'. Where relevant, please do not limit your reply to the situation at the time of submitting your response but provide details on any changes which may have occurred since 28 August 2023 or the date otherwise specified in the specific question.

Important considerations

Where you have already provided the requested information as part of previous requests sent by the Commission, please reference them in your answer and, where relevant, provide updated information starting from the date your previous reply was limited to. Please indicate clearly what has changed since the previous reply.

When replying to the questions, please do so separately for each recommender system within TikTok. When the answers are similar between different recommender systems, feel free to reference previously provided answers.

When replying to this RFI, please make sure to explicitly cite and reference all the aforementioned provided information to clearly establish its purpose and context.

I. As regards TikTok's compliance with Article 34 and 35 of Regulation (EU) 2022/2065, in particular concerning risks for electoral processes and civic discourse

1. In TikTok's 2024 risk assessment report, "*the impact of recommender systems on the content that users consume*" has been identified as a "*priority factor*" posing "*potential risks across many of its assessments*" ([5]). However, the report provides only limited information on how the provider of TikTok analysed and assessed whether and how the design or the functioning of its different recommender systems influences such risks. In this context, please provide the following information:
    a. Pursuant to Article 34(3) of Regulation (EU) 2022/2065, please provide all the risk assessment "Modules" underlying the 2024 risk assessment report.

    b. Pursuant to Article 34(3) of Regulation (EU) 2022/2065, please provide all supporting documents identifying, analysing and assessing whether and how all of TikTok's recommender systems and other relevant algorithmic systems influence risks to electoral processes, civic discourse and public security, illegal hate speech, gender-based violence, and fundamental rights, including the pluralism of the media.

([5]) See: "TikTok DSA risk assessment report 2024" submitted by the provider of TikTok to the European Commission on 29 August 2024, registered under Ares(2024)6149012, pp. 2 and 7.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014751

c. Please provide a copy of the presentation of the 2024 risk assessment report to the provider of TikTok's board of directors.

d. Please provide a non-technical description of the differences between the recommender systems currently deployed by the provider of TikTok and publicly available recommendation algorithms, architectures, and techniques, such as Monolith ([6]), CowClip ([7]), and HLLM ([8]).

2. In TikTok's 2024's risk assessment report, the provider of TikTok identifies that "*there is an inherent risk that features and advancements in content moderation models may be developed in a biased or discriminatory manner. This risk arises from the potential development of these features and policies using insufficiently inclusive data sets, safeguard and safety controls*" ([9]). Please explain whether the provider of TikTok analysed and assessed whether potentially biased content moderation systems may impact recommendations and the functioning of the specific recommender systems.

3. Search suggestions (e.g., 'Others searched for') on TikTok are becoming a relevant source of information, especially for young users, and particularly in the context of electoral processes. In this respect, recent research indicates how search suggestion systems employed on TikTok may amplify, among others, risks for electoral processes and civic discourse ([10]). Please provide the following information:
   a. How has TikTok identified, analysed and assessed the extent to which risks for elections and civic discourse are impacted by the design or functioning of its search suggestion systems?

   b. Which mitigation measures has TikTok put in place to mitigate risks for election and civic discourse that may arise from the design or functioning of its search suggestion systems?

   c. How does the provider of TikTok ensure that search suggestions across its platform direct users to authoritative sources?

([6]) See: Liu et al. (2022). Monolith: Real Time Recommendation System With Collisionless Embedding Table. In *Proceedings of 5th Workshop on Online Recommender Systems and User Modeling*, in conjunction with *the 16th ACM Conference on Recommender Systems (ORSUM@ACM RecSys 2022)*. ACM, New York, NY, USA.

([7]) See: Zheng et al. (2023). CowClip: Reducing CTR Prediction Model Training Time from 12 Hours to 10 Minutes on 1 GPU. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(9), 11390-11398. https://doi.org/10.1609/aaai.v37i9.26347, accessed on 26 September 2024.

([8]) See: Chen et al. (2024). HLLM: Enhancing Sequential Recommendations via Hierarchical Large Language Models for Item and User Modeling. *arXiv preprint*. https://arxiv.org/abs/2409.12740, accessed on 26 September 2024.

([9]) See: "TikTok DSA risk assessment report 2024" submitted by the provider of TikTok to the European Commission on 29 August 2024, registered under Ares(2024)6149012, p. 80.

([10]) For example, as reported by Interface and AI Forensic, TikTok's search suggestion feature "*suggests oddly specific and partially misleading search ideas that can lead users to believe questionable information or send them down contentious political rabbit holes*". See: https://tiktok-audit.com/blog/2024/Search-Suggestions/, accessed on 26 September 2024; See also: https://crossover.social/wp-content/uploads/2024/07/TheTiktokEffect.pdf, accessed on 26 September 2024.

d. How does TikTok apply content moderation on its search suggestion systems to mitigate risks for electoral processes and civic discourse, especially from users intentionally circumventing mitigation measures ([11])?

e. What is the accuracy of content moderation systems applied on TikTok's search suggestion systems across official EU languages, both in offline testing and online production environments?

4. Ahead of the last European elections, the provider of TikTok announced the introduction of labels on elections-related content. Labels would direct people to relevant Election Centres ([12]). In a post-election blogpost, the provider of TikTok stated that "*We actively directed people to these Centres through prompts on relevant election content and election-related searches, which were labelled with notice tags and search banners respectively*" ([13]). Recent research from Finland ([14]) showed an inconsistent approach to labelling of elections-related content in the months before the European elections. In this regard, please provide the information requested below for the different systems, features or sections in the platform's online interface, including the 'For You Feed', 'You May Like', 'LIVE', suggested accounts, suggested search, displayed ads, and banners:
   a. A detailed description and internal documents concerning how the labelling system related to elections-content is implemented by the provider of TikTok in the relevant period, including information on how recommendations are affected by these labels, as well as on any testing in place to assess the effectiveness of labelling as a mitigation measure;

   b. Internal documents detailing guidelines, policies, and practices to review, revise, and adapt the previously referenced labelled content, including any performance metrics or internal controls currently tracked that may trigger changes in these guidelines, policies, and practices, such as false negatives, false positives, or user complaints, among others;

   c. Information on whether any differences exist across EU Member States and official EU languages, including on testing outcomes, labelling strategies and effectiveness, as well as on the tracked performance metrics or internal controls.

5. In its 2023 risk assessment report, the provider of TikTok asserts that *"TikTok prohibits impersonation, which refers to accounts that pose as another person or*

([11]) For example, see: https://www.isdglobal.org/wp-content/uploads/2024/09/TikTok-White-Supremacy.pdf, accessed on 26 September 2024.
([12]) See: https://newsroom.tiktok.com/en-eu/our-work-to-prepare-for-the-2024-european-elections, accessed on 26 September 2024
([13]) See: https://newsroom.tiktok.com/en-eu/how-we-kept-our-community-safe-during-the-2024-european-parliament-elections, accessed on 26 September 2024.
([14]) See: https://crossover.social/wp-content/uploads/2024/07/TheTiktokEffect.pdf, accessed on 26 September 2024.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

entity in a deceptive manner. TikTok also reviews political accounts on the Platform in order to verify whether they are bona fide. TikTok does allow parody accounts" ([15]). TikTok's Help Center states, *"TikTok is an entertainment platform where millions of people come every day to watch, share, and create entertaining, joyful content. Some of this content includes current events in elections and political issues. Our priority in this space is to make sure we keep harmful misinformation off the platform and ensure our community has a positive and welcoming experience. That's why we classify relevant political TikTok accounts as a Government, Politician, and Political Party Account (GPPPA) and apply a number of policies that help prevent misuse of certain features*" ([16]). Recent research indicates a significant number of highly active accounts spreading political content on the platform, while having neither clear affiliations nor indications of authenticity ([17]). Please provide:

a. A detailed description of the internal review process for political accounts mentioned in the 2023 "Risk Assessment 2: Integrity & Misinformation Module 1: Risks to Election & Civic Integrity" ([18]);

b. A detailed description of the measures in place to detect political accounts on TikTok, including performance metrics or internal controls used to evaluate the detection effectiveness (e.g., precision, recall, false negatives);

c. A detailed description of the measures in place to detect inauthentic accounts or impersonation on TikTok, particularly for political accounts, including performance metrics or internal controls used to evaluate the detection effectiveness (e.g., precision, recall, false negatives);

d. A detailed description of how the provider of TikTok distinguishes between inauthentic accounts and parody accounts;

e. For Government, Politician, and Political Party Accounts (GPPPAs), clarify whether any recommender system rules, configurations or parameters affect their visibility on the platform;

f. Please explain all the measures in place for GPPPAs to mitigate any risk to electoral processes and civic discourse stemming from its specific recommender systems.

([15]) See: TikTok DSA risk assessment report, submitted by the provider of TikTok to the European Commission on 29 September 2023, registered under Ares (2023)6749619, p. 36.

([16]) See: https://support.tiktok.com/en/using-tiktok/growing-your-audience/government-politician-and-political-party-accounts, accessed on 26 September 2024.

([17]) See: https://democracy-reporting.org/en/office/EU/publications/inauthentic-behaviour-on-tiktok-concerning-accounts-support-the-afd-and-rassemblement-national, accessed on 26 September 2024.

([18]) See: Risk Assessment 2: Integrity & Misinformation. Module 1: Risks to Election & Civic Integrity, registered under Ares(2023)8221551, p. 36.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_014754

6. In its risk assessment reports, the provider of TikTok mentions 'borderline content' or 'borderline claims' as part of the Fact-Checking Programme, i.e. content that does not violate TikTok's community guidelines, but it is close to the threshold required to be considered as violative. In the context of 'Election Misinformation', please provide:
    a. A detailed explanation of how the specific recommender systems deal with 'borderline content' related to 'Election Misinformation'. Include information on any changes to the ranking, visibility, or prevalence of such content, such as after detection or after a certain threshold of content is reached;

    b. Specify whether any guidelines, policies, and practices are in place for the detection and monitoring of 'borderline content' related to 'Election Misinformation'. Include any performance metrics or internal controls used to track the increase, decrease, or general evolution of 'borderline content'. For all provided performance metrics or internal controls, provide the results of tests carried out in both offline and online production environments.

7. In its 2023 risk assessment, the provider of TikTok presents "*diversified content recommendations*" as a mitigation measure to prevent 'Election Misinformation' from being recommended or surfaced in a user's 'For You Feed': "*our recommendation system works to intersperse recommendations that might fall outside people's expressed preferences, offering an opportunity to discover new categories of content*" ([19]). Please provide with regards to "*political content*":
    a. Any internal guidelines, policies, and practices that may affect recommended content in terms of diversity or dispersion and any internal documents about testing performed "*to avoid recommending a series of similar content to protect against viewing too much of a content category that may be fine as a single video but potentially problematic if viewed in clusters*" ([20]);

    b. Detailed information on how such practices are assessed for effectiveness, specifying whether any additional measures are implemented in proximity to electoral processes, including information on how early they are activated.

8. In its 2024 risk assessment, the provider of TikTok states "*TikTok has deployed improved machine learning models to facilitate the effective detection of potential Election Misinformation. The primary detection model now takes into account relevant subpolicies for more accurate identification of violative content (e.g. Election Misinformation, and Medical Misinformation). This means that the model*

([19]) See: Risk Assessment 2: Integrity & Misinformation. Module 1: Risks to Election & Civic Integrity, p. 27, registered under Ares(2023)8221551.
([20]) See: Risk Assessment 2: Integrity & Misinformation. Module 1: Risks to Election & Civic Integrity, p. 28, registered under Ares(2023)8221551.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

now takes account of different possible types of misinformation and their respective varying characteristics, rather than purely basing its assessment on a binary outcome" ([21]). With regards to risks to electoral processes and civic discourse, please provide the following information:

a. An exhaustive list of input factors, features, signals, information, and metadata used by automated models to detect content that may contain 'Election Misinformation'. Include an explanation of how the provided list of input factors is modified to detect 'Election Misinformation' in a timely manner;

b. A description of how the importance of the listed input factors is measured and how it impacts a content's position, ranking, or order within the displayed recommendations;

c. A detailed description of when and how these machine learning models are used to detect potential 'Election Misinformation', specifying whether they are always active or only under certain circumstances, as well as on which type of content and platform surface they are used to detect potential 'Election Misinformation'.

9. Research indicates that TikTok users are exposed to conspiracy theories, illegal content, misinformation and disinformation related to electoral processes and civic discourse ([22]). In its risk assessment reports, the provider of TikTok mentions "*manual review*" of content as a mitigation measure against the risk of "*virality*" or the risks associated with "*viral content*". In its 2023 risk assessment report, the provider of TikTok states that "*[a]s video content rises in popularity in terms of the number of video views, it will be flagged for further manual review*" ([23]). Similarly, in its 2024 risk assessment report, the provider of TikTok states that it "*manually reviews all content that reaches a certain level of popularity (by monitoring the volume of video views)*" ([24]). Please provide:

a. A non-technical description of the detection mechanism(s) in place for the diverse types of content hosted on TikTok for "rises in popularity" and any other conditions for triggering a further manual review;

b. Clarification on the definition of "*virality*", detailing whether "*a certain level of popularity*" refers to specific threshold(s) in place, whether popularity varies by type of content, whether popularity is only considered in terms of number of views, or to which extent diffusion patterns are taken

([21]) See: "TikTok DSA risk assessment report 2024" submitted by the provider of TikTok to the European Commission on 29 August 2024, registered under Ares(2024)6149012 p. 66.

([22]) See: https://vsquare.org/far-right-tiktok-lies-conspiracies/ and https://www.isdglobal.org/digital_dispatches/naziok-an-organized-neo-nazi-tiktok-network-is-getting-millions-of-views/, accessed on 26 September 2024.

([23]) See: Risk Assessment 2: Integrity & Misinformation. Module 1: Risks to Election & Civic Integrity, registered under Ares(2023)822155, pp. 21 and 38.

([24]) See: "TikTok DSA risk assessment report 2024" submitted by the provider of TikTok to the European Commission on 29 August 2024, registered under Ares(2024)6149012, p.15.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

into account, including any other factors that may affect the impact that viral content may produce;

c. A description of the possible actions and outcomes from further manual reviews and, if applicable, a description of performance metrics or internal controls used to assess their effectiveness as risk mitigation measures;

d. If automated tools are used to detect "*viral content*" to be manually reviewed, please include:

   i. The respective technical system cards or reports, model cards or reports, datasheets or relevant internal documentation for such models, particularly for risk mitigation from the specific recommender systems;

   ii. A non-technical description of the training of such models, including a description of the training data used, how such data is selected and labelled, and which performance metric or loss function the models are optimizing for.

10. In its 2024 risk assessment report, TikTok mentions "*FYF eligibility standards to help ensure that any content promoted by its recommendation system is appropriate for a broad audience*" ([25]). However, clarification is required on the effectiveness of mitigating risks for civic discourse and electoral processes, and on whether eligibility standards also apply to other platform surfaces, such as 'You May Like', 'LIVE', suggested search, suggested accounts, displayed ads, banners, etc. In this context, please provide the following information:
    a. A detailed description of all the tests implemented to assess the effectiveness of the "FYF eligibility" standards as a mitigation measure against the risk to civic discourse, media diversity and pluralism, and electoral processes.

    b. All internal documents outlining the results of these tests, including their intended objective, goal, or purpose, as well as any performance metrics or internal controls used to evaluate, measure, or compare the test results. Include any thresholds used to define the results of these tests as effective or successful risk mitigation measures;

    c. A detailed description of any steps that have been taken to improve the results for future testing.

([25]) See: "TikTok DSA risk assessment report 2024" submitted by the provider of TikTok to the European Commission on 29 August 2024, registered under Ares(2024)6149012, p.87.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

11. The Guidelines for providers of VLOPs and VLOSEs on the mitigation of systemic risks for electoral processes ([26]), formally adopted on 26 April 2024, reference specific measures to mitigate risks posed by recommender systems in the context of electoral processes. In this regard, please reply to the questions below. In case you have already addressed the question above, please reference the specific answer it refers to.
    a. How has the provider of TikTok ensured that the specific recommender systems are designed and adjusted in a way that gives users meaningful choices and controls over their feeds, with due regard to media diversity and pluralism? If the provider of TikTok does not employ such mitigation measures, describe how it ensures media diversity and pluralism in its recommender systems and how it mitigates the risk of echo chambers caused by its recommender systems. In the latter case, please provide a timeline for the implementation of the measure.

    b. Which measures has the provider of TikTok established to reduce the prominence of disinformation in the context of elections based on clear and transparent methods? Which factors are used to determine whether such content classifies as disinformation in the context of elections, and how is such assessment made and by whom?

    c. Has the provider of TikTok adopted measures to reduce the prominence of disinformation in the context of elections, specifically coming from accounts that have repeatedly been found to spread disinformation? If yes, which ones?

    d. Has TikTok regularly assessed the performance and impact of its specific recommender systems and addressed any emerging risks or issues related to electoral processes? If yes, how?

    e. Which measures has the provider of TikTok established to provide transparency around the design and functioning of its specific recommender systems, in relation to the data and information used in designing systems that foster media pluralism and diversity of content, to facilitate third party scrutiny and research?

    f. Has the provider of TikTok engaged with external parties to conduct adversarial testing and red team exercises on these systems to identify potential risks such as risks stemming from biases, susceptibility to manipulation, or amplification of misinformation, disinformation, Foreign Information Manipulation and Interference, or other harmful content. In the case such testing has taken place, what were the results and how were learnings implemented? If yes, how?

([26]) See: Commission Guidelines for providers of Very Large Online Platforms and Very Large Online Search Engines on the mitigation of systemic risks for electoral processes pursuant to Article 35(3) of Regulation (EU) 2022/2065, OJ C/2024/3014, 26.04.2024, pp. 1-19.

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

ANNEX II

INSTRUCTIONS FOR REPLYING

The reply and relevant documents submitted under Regulation (EU) 2022/2065 shall be submitted in a format allowing the Commission to process them electronically and enabling their digitisation and character recognition. The reply must clearly indicate the number of the file and the date. The reply should be page-numbered and the paragraphs numbered.

Please submit your reply by sending it to the functional mailbox of the Directorate General for Communications Networks, Content and Technology at: CNECT-DSA-Registry@ec.europa.eu.

The electronic version of your reply shall be considered by the Commission as an original reply. There is therefore no need to submit a hard-copy in addition to the electronic version of the reply.

Electronically signed on 02/10/2024 10:17 (UTC+02) in accordance with Article 11 of Commission Decision (EU) 2021/2121

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED
TT_HJC_014759

Exhibit 297

Schedule of Annexes

The below Schedule contains a list of documents identified as responsive to the requests in the request for information Ref.Ares(2024)8523013 - 29/11/2024 (the "**RFI**"). TikTok further notes that:

- There are three responsive documents identified as being partially privileged. Therefore, for the purposes of protecting such privilege in other jurisdictions and contexts, and on the advice of external counsel, TikTok objects to the disclosure of such privileged information. We are available at your convenience to discuss.
- The documents listed in this Schedule necessarily reflect matters at the time the documents were created (or last updated, where applicable). As such, some of the information they contain may not fully reflect the current position. Accordingly, please refer to TikTok's responses to the questions of this RFI for the position as at the date of submission of those responses.
- Any documents listed in this Schedule have been redacted to remove the following information: (i) personal information; (ii) information that clearly relates solely to countries outside the EEA and is confidential and commercially sensitive; and/or (iii) legally privileged information.
- TikTok is producing documents under applicable laws but not waiving privilege claims that may be applicable in any jurisdictions. In the event that any waiver of privilege does nonetheless occur, such waiver is limited to the disclosure of the relevant document(s) to the Commission only and only for the purposes of TikTok's response to this RFI.

Question No.	Annex No.	Document	Privileged	Confidential	Redacted
1	003	**Re: [External] Reporting electoral promotion - TikTok**	No	Yes	Yes
1	004	**Re: [External] RE: check the use of bot networks in platforms**	No	Yes	Yes
1	005	**[External] Decision no. 39/22.11.2024 pronounced by the ELECTORAL DISTRICT OFFICE FOR ROMANIAN CITIZENS WITH DOMICILE OR RESIDENCE OUTSIDE THE COUNTRY NO. 43**	No	Yes	Yes
1	006	Information included in Annex 005	No	Yes	Yes
1	007	**[External] RE: Decision of the Electoral Bureau of County Constituency No. 3 Argeș – election of the Senate and the Chamber of Deputies in 2024 with no. 54/21.11.2024**	No	Yes	Yes
1	008	Information included in Annex 007 and 014	No	Yes	Yes
1	009	**[External] FW: Decision of the Electoral Bureau of County Constituency No. 3 Argeș – election of the Senate and the**	No	Yes	Yes

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Question No.	Annex No.	Document	Privileged	Confidential	Redacted
		Chamber of Deputies in 2024 with no. 54/21.11.2024			
1	010	Information included in Annex 009A	No	Yes	Yes
1	011	Information included in Annex 009B	No	Yes	Yes
1	012	**RE: [External] RE: check the use of bot networks in platforms**	No	Yes	Yes
1	013	**Re: [External] RE: check the use of bot networks in platforms**	No	Yes	Yes
1	014	**[External] RE: check the use of bot networks in platforms**	No	Yes	Yes
1	015	**Re: [External] RE: check the use of bot networks in platforms**	No	Yes	Yes
1	016	**Re: [External] RE: check the use of bot networks in platforms**	No	Yes	Yes
1	017	**Re: [External] RE: check the use of bot networks in platforms**	No	Yes	Yes
1	018	**[External] RE: check the use of bot networks in platforms**	No	Yes	Yes
1	019	**Re: [External] Electoral offenses, Violation of the confidentiality of the vote, Failure to comply with the ballot box regime**	No	Yes	Yes
1	020	**Re: [External] Electoral offenses, Violation of the confidentiality of the vote, Failure to comply with the ballot box regime**	No	Yes	Yes
1	021	**[External] Ministry of Research, Innovation and Digitalization request for urgent content removal**	No	Yes	Yes
1	022	Information included in Annex 021A	No	Yes	Yes
1	023	Information included in Annex 021B	No	Yes	Yes
1	024	**[External] Election**	No	Yes	Yes
1	025	**Re: [External] Election**	No	Yes	Yes
6	026	**RO Speaker Series.mp4**	No	Yes	No

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Question No.	Annex No.	Document	Privileged	Confidential	Redacted
6	027	**RO-party-GPPPA.pdf**	No	Yes	Yes
6	028	**[External] Invitation: RO Workshop United Right Alliance @ Thu Nov 7, 2024 5pm - 6pm (GMT+1) (Redacted @bytedance.com)**	No	Yes	Yes
6	029	Information included in Annex 028A	No	Yes	Yes
6	030	Information included in Annex 028B and 031	No	Yes	No
6	031	**[External] Invitation: RO Workshop USR @ Tue Nov 5, 2024 12pm - 1pm (GMT+1) (Redacted @bytedance.com)**	No	Yes	Yes
6	032	Information included in Annex 031	No	Yes	Yes
6	033	**Fw: [External] Re: counteracting disinformation campaign - October 2024**	No	Yes	Yes
6	034	**Re: [External] Re: counteracting disinformation campaign - October 2024**	No	Yes	Yes
6	035	**[External] Re: TikTok adcredits**	No	Yes	Yes
6	036	**Re: [External] Re: counteracting disinformation campaign - October 2024**	No	Yes	Yes
6	037	**Re: [External] Re: counteracting disinformation campaign - October 2024**	No	Yes	Yes
6	038	Information included in Annex 037	No	Yes	Yes
6	039	**Re: [External] Re: TikTok Videos - Funky Citizens**	No	Yes	Yes
6	040	**[External] Accepted: TikTok EU Elections Speaker Series Romania & Bulgaria @ Wed May 22, 2024 2pm - 3pm (IST) (Redacted bytedance.com)**	No	Yes	Yes
6	041	Information included in Annex 040	No	Yes	Yes
8	042	**[T&S] [Global/US] [Escalation-only] Paid Political Content Policy**	No	Yes	Yes
8	043	Information included in Annex 042A	No	Yes	No
8	044	Information included in Annex 042B	No	Yes	No

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Question No.	Annex No.	Document	Privileged	Confidential	Redacted
8	045	Information included in Annex 042C	No	Yes	No
8	046	Information included in Annex 042D	No	Yes	No
8	047	Information included in Annex 042E	No	Yes	No
8	048	Information included in Annex 042F	No	Yes	No
8	049	Information included in Annex 042G	No	Yes	No
8	050	Information included in Annex 042H	No	Yes	No
8	051	Information included in Annex 042I	No	Yes	Yes
8	052	Information included in Annex 042J	No	Yes	No
8	053	Information included in Annex 042K	No	Yes	No
8	054	**BI political model**	No	Yes	No
8	055	**[Moderation Guidance] - Political Content (Global)**	No	Yes	Yes
8	056	Information included in Annex 055A	No	Yes	No
8	057	Information included in Annex 055B	No	Yes	No
8	058	Information included in Annex 055C	No	Yes	No
8	059	Information included in Annex 055D	No	Yes	Yes
8	060	Information included in Annex 055E	No	Yes	Yes
8	061	Information included in Annex 055F	No	Yes	No
8	062	Information included in Annex 055G	No	Yes	No
8	063	Information included in Annex 055H	No	Yes	No
8	064	Information included in Annex 055I	No	Yes	No
8	065	Information included in Annex 055J	No	Yes	No
8	066	Information included in Annex 055K	No	Yes	No
8	067	Information included in Annex 055L	No	Yes	No
8	068	Information included in Annex 055M	No	Yes	No
8	069	Information included in Annex 055N	No	Yes	No
8	070	Information included in Annex 055O	No	Yes	No

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

TT_HJC_013499

Question No.	Annex No.	Document	Privileged	Confidential	Redacted
8	071	Information included in Annex 055P	No	Yes	No
8	072	Information included in Annex 055Q	No	Yes	No
8	073	Information included in Annex 055R	No	Yes	No
8	074	Information included in Annex 055S	No	Yes	No
8	075	Information included in Annex 055T	No	Yes	No
8	076	Information included in Annex 055U	No	Yes	No
8	077	Information included in Annex 055V	No	Yes	No
8	078	Information included in Annex 055W	No	Yes	No
8	079	Information included in Annex 055X	No	Yes	No
8	080	Information included in Annex 055Y	No	Yes	No
8	081	Information included in Annex 055Z	No	Yes	No
8	082	Information included in Annex 055AA	No	Yes	No
8	083	Information included in Annex 055BB	No	Yes	No
8	084	Information included in Annex 055CC	No	Yes	No
8	085	Information included in Annex 055DD	No	Yes	No
8	086	Information included in Annex 055EE	No	Yes	No
8	087	Information included in Annex 055FF	No	Yes	No
8	088	**Global SOP - Branded Content**	No	Yes	Yes
8	089	Information included in Annex 088A	No	Yes	No
8	090	Information included in Annex 088B	No	Yes	No
8	091	Information included in Annex 088C	No	Yes	No
8	092	Information included in Annex 088D	No	Yes	Yes
8	093	Information included in Annex 088E	No	Yes	No
8	094	Information included in Annex 088F	No	Yes	No
8	095	Information included in Annex 088G	No	Yes	No
8	096	Information included in Annex 088H	No	Yes	No

Question No.	Annex No.	Document	Privileged	Confidential	Redacted
8	097	Information included in Annex 088I	No	Yes	No
8	098	Information included in Annex 088J	No	Yes	No
8	099	Information included in Annex 088K	No	Yes	No
8	100	**[T&S][M10N Policy] Undisclosed Branded Content Policy - XFN Sharing**	No	Yes	Yes
8	101	Information included in Annex 100A	No	Yes	No
8	102	Information included in Annex 100B	No	Yes	No
8	103	Information included in Annex 100C	No	Yes	No
8	104	Information included in Annex 100D	No	Yes	No
8	105	Information included in Annex 100E	No	Yes	No
8	106	Information included in Annex 100F	No	Yes	No
8	107	**[T&S] Branded Content video moderation process//Brand Content** 视频审核流程	**Partially**	Yes	Yes
8	108	Information included in Annex 107A	No	Yes	No
8	109	Information included in Annex 107B	No	Yes	Yes
8	110	Information included in Annex 107C	No	Yes	Yes
8	111	**(P&C) [T&S][Video/LIVE][Global][Escalation Only] Undisclosed Branded Content Policy Proposal**	No	**No**	Yes
8	112	Information included in Annex 111A	No	Yes	No
8	113	Information included in Annex 111B	No	Yes	No
8	114	Information included in Annex 111C	No	Yes	No
8	115	Information included in Annex 111D	No	Yes	No
9	116	**CRI Assessment - Romania's Presidential and Parliamentary Elections**	No	Yes	Yes
9	117	Information included in Annex 116	No	Yes	No
9	118	**[IOTD] - Democracy Reporting Triage**	No	Yes	Yes

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Question No.	Annex No.	Document	Privileged	Confidential	Redacted
9	119	**IOID Operational Assessment: Influence Operations Targeting Romania**	No	Yes	Yes
9	120	**IOID Romanian Election Lead: Blunted Impale**	No	Yes	Yes
9	121	**IOID Romanian Election Lead: Meddling Vines**	No	Yes	Yes
9	122	**IOID Romanian Election Lead: Wallachian Wave**	No	Yes	Yes
9	123	Information included in Annex 122	No	Yes	Yes
9	124	**[G-IOTD] Disruption Form - Phony Praise**	No	Yes	Yes
9	125	Information included in Annex 124A	No	Yes	Yes
9	126	Information included in Annex 124B	No	Yes	No
9	127	Information included in Annex 124C	No	Yes	No
9	128	Information included in Annex 124D	No	Yes	Yes
9	129	Information included in Annex 124E	No	Yes	No
9	130	Information included in Annex 124F	No	Yes	Yes
9	131	Information included in Annex 124G	No	Yes	No
9	132	Information included in Annex 124H	No	Yes	No
9	133	Information included in Annex 124I	No	Yes	Yes
9	134	Information included in Annex 124J	No	Yes	No
9	135	Information included in Annex 124K	No	Yes	Yes
9	136	**[G-IOTD] LIVE Interference - Investigation**	No	Yes	Yes
9	137	Information included in Annex 136A	No	Yes	Yes
9	138	Information included in Annex 136B	No	Yes	Yes
9	139	**[G-IOTD] Disruption Form - Bot Race**	No	Yes	Yes
9	140	Information included in Annex 139A	No	Yes	Yes
9	141	Information included in Annex 139B	No	Yes	No

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Question No.	Annex No.	Document	Privileged	Confidential	Redacted
9	142	Information included in Annex 139C	No	Yes	No
9	143	Information included in Annex 139D	No	Yes	Yes
9	144	Information included in Annex 139E	No	Yes	No
9	145	Information included in Annex 139F	No	Yes	Yes
9	146	Information included in Annex 139G	No	Yes	Yes
9	147	Information included in Annex 139H	No	Yes	Yes
9	148	Information included in Annex 139I	No	Yes	No
9	149	Information included in Annex 139J	No	Yes	No
9	150	Information included in Annex 139K	No	Yes	No
9	151	Information included in Annex 139L	No	Yes	No
9	152	**STR Investigation - Romanian Information Landscape Analysis**	No	Yes	Yes
9	153	**IOID RFI: #EchilibruSiVerticalitate**	No	Yes	Yes
9	154	**GSO IO - Status Report - Romanian Elections - 11/27/24**	No	Yes	Yes
9	155	**XFN Inauthentic Accounts/ BOT Nets Investigation (31 Oct) - Romanian Elections 2024**	No	Yes	Yes
9	156	**[G-IOTD] Disruption Form - Presidential Propagate**	No	Yes	Yes
9	157	Information included in Annex 156A	No	Yes	No
9	158	Information included in Annex 156B	No	Yes	No
9	159	Information included in Annex 156C	No	Yes	No
9	160	Information included in Annex 156D	No	Yes	No
9	161	Information included in Annex 156E	No	Yes	No
9	162	Information included in Annex 156F	No	Yes	No
9	163	Information included in Annex 156G	No	Yes	No
9	164	**[G-IOTD] Disruption Form - AI Author**	No	Yes	Yes
9	165	Information included in Annex 164A	No	Yes	No

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Question No.	Annex No.	Document	Privileged	Confidential	Redacted
9	166	Information included in Annex 164B	No	Yes	No
9	167	Information included in Annex 164C	No	Yes	No
9	168	Information included in Annex 164D	No	Yes	No
9	169	Information included in Annex 164E	No	Yes	Yes
9	170	Information included in Annex 164F	No	Yes	Yes
9	171	**[IOTD] #ESV Triage**	No	Yes	Yes
9	172	Information included in Annex 171	No	Yes	Yes
9	173	**IOID Tipper: Georgescu for President**	No	Yes	Yes
9	174	Information included in Annex 173	No	Yes	Yes
9	175	**IOID Lead: Live Show**	No	Yes	Yes
9	176	Information included in Annex 175	No	Yes	Yes
9	177	**IOID Tipper: Bot Race**	No	Yes	Yes
12	178	**RO Election candidates related hashtags**	No	Yes	Yes
12	179	**[T&S] Keywords Tracker | Romanian Elections 2024**	No	Yes	Yes
12	180	**[T&S] Romanian Elections 2024 - Control Tower**	No	Yes	Yes
12	181	**Romanian Elections**	**Partially**	Yes	Yes
12	182	**[RRS] RO Election Report_Tik_Tok**	**Partially**	Yes	Yes
12; 14	183	**RO elections; proactive & esc support**	No	Yes	Yes
12; 14	184	**RO elections; proactive & esc support**	No	Yes	Yes
14	185	**[T&S] Elections Metric Framework and Definitions**	No	Yes	Yes
14	186	Information included in Annex 185A	No	Yes	Yes
14	187	Information included in Annex 185B	No	Yes	No
14	188	Information included in Annex 185C	No	Yes	No
14	189	Information included in Annex 185D	No	Yes	No
14	190	Information included in Annex 185E	No	Yes	No

CONTAINS BUSINESS CONFIDENTIAL INFORMATION.
CONFIDENTIAL TREATMENT REQUESTED

Question No.	Annex No.	Document	Privileged	Confidential	Redacted
14	191	Information included in Annex 185F	No	Yes	No
14	192	**George Simion appeal**	No	Yes	Yes
14	193	Information included in Annex 192	No	Yes	No

***

Exhibit 298

Bing

Bing Systemic Risk Assessment Report

August 2025

Microsoft

MICROSOFT CONFIDENTIAL MSFT_JUD_000002441

Contents

Executive Summary 4
Introduction 6
Purpose 6
Document Structure 6
Bing's Approach to Protecting Users and their Fundamental Rights Online 8
Bing's Values and Commitments 8
Bing's Approach to User Safety 9
Generative AI Experiences Specific Mitigations 19
How Bing monitors effectiveness of risk mitigation efforts 22
Bing Overview and Risk Profile 23
User Base and Market 23
Considerations for Bing's Risk Profile 23
Core Search Features 24
Generative AI Experiences 28
Ancillary Search Features Risk Profile 35
Changes to Bing Features 37
Systemic Risk Assessment Methodology 39
Methodology Overview 39
Risk Assessment Process 39
Assessment Tooling 42
Risk Assessment Results 43
Overall Results 43
Changes to Bing's Risk Profile 44
Emerging Trends and Shifts in Global Risk Posture 44
Changes in Risk Rating 45
Results per Risk Area 47
Conclusion 89
Conclusion of Assessment 89
Progress on Enhanced Mitigations from Bing's Last Report 91
Looking Ahead to the Coming Year 92
Appendix I: Detailed Systemic Risk Assessment Methodology 94
Risk Areas 94
Risk Factors and Influencers 95
Risk Assessment Inputs 95
Probability 96
Severity 97
Inherent Risk 98

Maturity of Mitigations 99
Residual Risk 101
Sample scoring calculation 102
Appendix II: Bing's Digital Safety Risk and Compliance Framework and Mitigations 103
Product Quality 103
Product Safety 106
Targeted Interventions 112
Content Moderation 115
Incident Response 119
Research and Partnership 121
Transparency and Reporting 123
Metrics x Funnel 125
Appendix III: Terms and Acronyms 128
Appendix IV: Summary of External Consultations and Partner Insights 131

Executive Summary

In a technology-driven world, search engines are in many ways the gateway to the Internet and the primary way people find the content they are looking for amongst the trillions of ever-changing webpages available online. Search engines, including Bing, play a vital role in upholding the fundamental rights of free and open access to information and free expression. At the same time, Bing recognizes that other fundamental rights and social interests, such as privacy, safety, democratic processes, public health, and national security, are also vital to Bing users and to a healthy society, and Bing must balance these interests and rights to maximize benefits to users while minimizing possible harms.

Addressing content risks in a search service often requires a different approach than may be appropriate for other types of online services. Over-moderation of content in search could have a significant negative impact on the right to access information, freedom of expression, and media plurality. Therefore, Bing must carefully balance these competing fundamental rights and interests as it works to ensure its algorithms return the highest quality content relevant to a user's queries without unduly limiting its ability to access answers to the questions they ask.

Bing has implemented a robust and multi-layered approach of five functions and related measures to manage risks and maintain this careful balance. This approach and the implemented mitigation measures are anchored in the principles of:

- developing ***fair, reliable, safe, private, secure, inclusive, transparent,*** and ***accountable*** algorithmic systems;
- providing ***credible and authoritative results*** relevant to user queries;
- ***promoting free and open access to information*** within the bounds of the law and with respect for local law and other fundamental rights, such as privacy and public safety;
- taking steps to ***protect users from harmful and unexpected offensive content***; and
- ***being transparent*** with users about Bing's principles and practices, as well as Bing's decisions and actions.

The measures address systemic risk in various ways, including but not limited to: algorithmically ranking content, grounding results in high authority content, removing content when necessary, managing incident response, and enabling user feedback and reports, in order to provide a safe experience for all users. Teams across Bing continuously measure the effectiveness of the risk mitigation measures implemented in alignment with these principles to adjust and refine them as threats and risks evolve.

Bing is also continuously innovating and enhancing its service to help users find exactly what they're looking for faster. Bing's innovation is rooted in its values and respect for user fundamental rights and safety, and Bing considers – throughout the innovation process – the potential for product changes to reduce or increase the systemic risk profile of the service.

Finally, Bing as a service does not exist in a vacuum but within an ecosystem of geopolitical and cultural trends that influence risk vectors. Bing tracks these trends as part of its continuous monitoring to align resources as needed to dynamically address risks. For example, there has been growing concern about the risks posed by GenAI, particularly regarding Non-Consensual Intimate Imagery ("NCII"), Concern about violent extremism also remains high, although with changes in specific actors and targets. Bing uses information on trends to inform product and policy priorities and takes targeted actions consistent with the principles above and described further at How Bing delivers search results.

Continuous monitoring of the ecosystem as well as Bing usage trends is necessary and key to keeping Bing services safe for all users. Annual assessments are also useful, to reflect on trends over discrete periods of time to evaluate whether adjustments in risk mitigation measures are warranted to address more gradual shifts in risk. Therefore, in addition to continuous monitoring activities, Bing conducts an annual assessment of the effectiveness, proportionality, and reasonableness of the implemented mitigation measures relative to potential systemic risks stemming from the use, misuse, or functioning of the service pursuant to Article 34 of the EU Digital Services Act ("DSA").

The results of Bing's continuous monitoring and annual Systemic Risk Assessment, factoring in external shifts and internal product changes, indicate that residual risks to users, stemming from use, misuse, or functioning of Bing services, remains consistent and primarily falls within the Low range, with a few exceptions. Bing assesses that the mitigation measures in place across all risk areas are proportionate, reasonable, and effective to reduce risks on the service to an acceptable level, based on implementation of best available scientific and technical insights and the incorporation of internal and external data and expert opinion.

In the 2025 Systemic Risk Assessment, Bing assessed eight of the 12 risk areas as Low Residual Risk. Of the four remaining risk areas, Freedom of Expression and Information is rated as Minimal Residual Risk, while Human Dignity, Illegal Content and Activities, and Private and Family Life are rated as Moderate Residual Risk. The Illegal Content and Activities Residual Risk Rating increased to Moderate solely due to the clerical internal reassignment of risks related to Child Sexual and Abuse Material ("CSAM") from the Rights and Protection of Minors risk area to the Illegal Content and Activities risk area. Based on Bing's unique risk profile as a search engine, the residual risks for the Human Dignity, Illegal Content and Activities, and Private and Family Life areas are within an acceptable range; at the same time, these challenges are evolving and require ongoing monitoring and investment.

Despite the relatively low risk overall, Bing is committed to continually enhancing user trust and safety and, therefore, as part of both this and last year's processes, identified opportunities to further enhance the systemic risk mitigation measures currently in place.

Bing drove progress in each of the six areas highlighted in the 2024 Systemic Risk Assessment report; for example, Bing launched a dashboard showing its core safety metric, the Defensive Defect Rate, aligned to DSA risk areas, for more tailored analysis, monitoring, and reporting. As part of this 2025 Systemic Risk Assessment, Bing has identified the following five opportunities to enhance user trust and safety:

1) Enhancing governance and pre-launch review.
2) Developing and refining policies to align user experience with Bing's Trustworthy Search Principles.
3) Continuing investments in targeted interventions to better manage systemic risks.
4) Enabling greater data-driven improvement by refining documentation, resources, and metrics.
5) Improving children's safety.

Bing will work to implement these enhanced mitigations and will continue to monitor evolving risks on the service to enhance and refine measures as needed to maintain Bing's overall Low risk profile.

Introduction

Purpose

This Systemic Risk Assessment Report ("the Report") is responsive to the obligations listed in Article 42.4 (a) and (b) of regulation (EU) 2022/2065, known as the Digital Services Act (DSA), in setting out the results of the Systemic Risk Assessment, which was conducted pursuant to Article 34, and the specific mitigation measures put in place pursuant to Article 35 (1).

The Report addresses the Bing service, which was designated by the European Commission on 25 April 2023 as a Very Large Online Search Engine ("VLOSE"). The Bing service includes the services offered on Bing.com, Bing mobile apps, and relevant features, as described in the section Bing Overview. This includes core search features; ancillary services, including search verticals such as images, shopping, maps, news, travel, and real estate, and advertisements in Bing; and the GenAI-enhanced experiences, Copilot Search, Bing Image Creator, and Bing Video Creator.

The report describes the assessment of systemic risks on the Bing VLOSE impacting the EU for the assessment period of August 2024-July 2025.

The report discusses key measures implemented to mitigate these risks; the evaluation of their reasonableness, effectiveness, and proportionality to the identified risks; and areas for continued improvement of mitigations.

Document Structure

This document is structured to provide a comprehensive overview of Bing's user protection measures, its services, and its approach to systemic risk management under the DSA. It begins with a description of the safeguards in place to uphold users' fundamental rights online, followed by a description of Bing, its key features, and its risk profile. The document then outlines the methodology used to identify and evaluate systemic risks across the Bing service. This includes an assessment of how such risks may manifest in practice and how existing mitigations address those risks. Finally, it details Bing's strategy for managing Residual Risks, including prioritization of mitigations based on the risk evaluation outcomes.

The structure is designed to first establish an understanding of the service and its safety framework, and then to systematically examine the alignment between identified risks and the effectiveness of protections intended to mitigate potential harm to users and fundamental rights.

This Report includes the following components:

- **Bing's Approach to Protecting Users and their Fundamental Rights Online:** An overview of (1) the principles Bing follows to provide users with high quality, effective, and safe services coupled with its approach to protecting fundamental rights and (2) Bing's approach to user safety through a multi-layered approach.
- **Bing Overview and Risk Profile:** An overview of (1) the Bing service and its key features across three categories (core search, GenAI experiences, and ancillary search features) as well as updates to Bing features during the assessment period and (2) Bing's risk profile and key considerations due to its nature as a search engine.

- **Systemic Risk Assessment Methodology:** An overview of the process Bing used to conduct this year's Systemic Risk Assessment including an overview of the methodology used, a description of the risk assessment process, and an explanation of the assessment tools used. Further detail, including risk scoring definitions, criteria, and rating scales, is included in Appendix I: Detailed Risk Assessment Methodology.
- **Risk Assessment Results:** A description of (1) each risk area, the risk analysis conducted, and the key relevant implemented mitigations, (2) an explanation of changes to Bing's risk profile, (3) emerging trends and shifts in global risk program, (4) changes in risk ratings, and (5) a deep dive into the results per risk area.
- **Conclusion:** A conclusion of the assessment, a description of progress made on enhanced mitigations from Bing's last report, and a description of the heightened focus areas for the next assessment period.
- **Appendix I - Detailed Systemic Risk Assessment Methodology:** Detailed explanation of the methodology used for this year's Systemic Risk Assessment, including the risk definitions, inputs, risk scoring criteria, and rating scales.
- **Appendix II – Bing's Digital Safety Risk and Compliance Framework and Mitigations:** An overview of the mitigations that apply across the Bing service and across DSA systemic risks.
- **Appendix III – Terms and Acronyms:** Definitions of the acronyms and initialisms mentioned throughout the report.
- **Appendix IV – Summary of External Consultations and Partner Insights:** A synopsis of Bing's engagement with external parties.

Bing's Approach to Protecting Users and their Fundamental Rights Online

Bing's Values and Commitments

Given that most research today is conducted online, search engines play a vital role in society by promoting fundamental rights to Freedom of Expression and Information and supporting media pluralism. It is integral to Bing's design principles to ensure that Bing does not unduly restrict users' ability to search and receive information. At the same time, Bing recognizes that other fundamental rights and social interests, such as privacy, safety, upholding democracy, public health, and national security, are also vital to Bing's users and to a healthy society, and Bing must balance these interests and rights to maximize benefit to users while minimizing possible harms.

In order to provide its users with a high quality, effective, and safe search service that appropriately balances tradeoffs between fundamental rights, Bing follows the below "*Trustworthy Search Principles*" to guide the product design, experience, algorithms, and mitigation measures that Bing adopts to ensure users' expectations are met while addressing potential risks or harms arising from use of the service. These principles also underpin Bing's GenAI experiences, which represents a very small portion of users' experiences. Bing is evolving its policies, operations, and infrastructure to better identify and manage the associated GenAI risks.

Bing aims to provide credible and authoritative results relevant to user queries.

- Bing works to provide the highest quality, authoritative content relevant to users' queries (further detail on high quality, authoritative ranking under Product Quality).
- Bing's goal is to provide fair, balanced, and comprehensive content. When there are multiple credible perspectives, Bing tries to display them in informative ways.
- When there is no authoritative source, Bing's goal is to avoid promoting bias or potentially misleading information.
- Bing respects user intent. When a user expresses a clear intent to access specific information, Bing provides relevant results even if they are less credible, while (as described in more detail below) working to ensure that users are not misled by such search results.

Bing promotes free and open access to information within the bounds of the law and with respect for local law and other fundamental rights, such as privacy and public safety.

- Bing provides open access to as much of the web as possible, but in limited cases it may undertake certain interventions (such as removal of a website or downranking) for instances where the content violates local law or Microsoft's policies.
- When limiting access to content, Bing strives to ensure its actions are narrowly tailored, so it does not unduly restrict important interests, such as the freedom of expression, open access to information, and media pluralism, and provides transparency regarding its actions.

Bing takes steps to protect users from harmful and unexpected offensive content.

- Bing recognizes that there are many reasons why someone may want to research or review harmful or controversial content but also recognizes the importance of ensuring users are not inadvertently misled by or unintentionally shown such content.

- For certain types of content where Bing identifies results that may include harmful or misleading information, Bing may provide supplemental information, such as warnings and public service announcements ("PSAs"), to inform users about potential risks.
- Bing gives users control over the type of content they encounter in Bing through SafeSearch.
- Absent a clear intent to access specific content, Bing assumes an intent to find high authority results.

Bing is transparent about its principles and practices, as well as its decisions and actions.

- Bing provides users with information about its principles regarding ranking and relevance, and moderation policies.
- When Bing limits access to content, where relevant Bing provides notice to users that content was removed.
- Bing provides information about content moderation actionstaken in DSA transparency reports.

AI systems in Bing are developed and evaluated in accordance with Microsoft's Responsible AI ("RAI") Standard:

- Fairness: How might an AI system allocate opportunities, resources, or information in ways that are fair to the humans who use it?
- Reliability and Safety: How might the system function well for people across different use conditions and contexts, including ones it was not originally intended for?
- Privacy and Security: How might the system be designed to support privacy and security?
- Inclusiveness: How might the system be designed to be inclusive of people of any ability?
- Transparency: How might people misunderstand, misuse, or incorrectly estimate the capabilities of the system?
- Accountability: How can Microsoft create oversight so that humans are accountable and in control?

Additional information is available at How Bing Delivers Search Results and in Bing's Webmaster Guidelines.

In addition, Bing is guided by Microsoft's broader corporate mission and commitments to:

- Expand Opportunity
- Earn Trust
- Protect Fundamental Rights
- Advance Sustainability

More information on Microsoft's broader commitment to responsible practices, including sustainability, ethics, and human rights, is available on Microsoft's Corporate Social Responsibility Report.

Bing's Approach to User Safety

Microsoft respects freedom of expression and the right to access information. At the same time, in accordance with Microsoft principles, policies, and standards for RAI, privacy, digital safety, information integrity and the Bing Trustworthy Search Principles, Bing has implemented a robust and multi-layered system of mitigations to manage risks and maintain this careful balance. These measures are organized into five layers and represented in the Bing Safety Funnel (Figure 1)

The Bing Safety Funnel is applied globally and progresses through five layers:

1. Product Quality;

2. Product Safety;
3. Targeted Interventions;
4. Content Moderation; and
5. Incident Response.

Each layer incorporates specific standards, processes, and mitigations that apply across the Bing ecosystem. As illustrated by the funnel shape, interventions are applied to an increasingly narrower set of queries with highly specific intent. This layered system provides a structured defense that helps users safely find the relevant and trustworthy information they are seeking.

Underpinning the Bing Safety Funnel is Bing's strong emphasis on data-driven safety governance. Specifically, Bing relies on metrics to measure the real-world performance of its safety systems and to guide ongoing improvements. Central to this approach is the Defensive Defect Rate ("DDR"), Bing's core safety metric, which is used to measure the presence of content that violates Bing policies and the efficacy of implemented safety. DDR allows Bing to evaluate the effectiveness of its mitigations, identify regressions, and take corrective action. In addition to overall DDR, Bing now also tracks DDR in the DSA systemic risk areas, enabling more tailored monitoring, analysis, and reporting. Bing reviews DDR regularly in both pre-launch assessments and in production settings. This supports accountability and drives system-wide refinement.

A visualization of the Bing Safety Funnel, with its five layers of safety mitigations working together as a system, and underpinned with metrics, is provided below. An inventory of Bing's service-wide mitigations, organized by the layers of the Bing Safety Funnel, is included in Appendix II: Bing's Digital Safety Risk and Compliance Framework Mitigations.

Figure 1: Bing Safety Funnel

The following sections first highlight *select* service-wide mitigations within each layer of the Bing Safety Funnel to illustrate the broad purpose of the layer and then include the *complete* set of GenAI-specific mitigations applied across the service; mitigation titles are in bold and italic font. Then, they explain how Bing

continuously measures, monitors the effectiveness of this layered approach to safety, and uses metrics to drive continuous improvement. Bing's service-wide mitigations, found in their entirety in Appendix II, broadly apply to all products and features, with specific feature-level mitigations applied in addition. This layered system, based on and continuously monitored through metrics, enables Bing to proactively manage risks and respond to real-world use cases.

Product Quality

At the top of the Bing Safety Funnel is the Product Quality layer. This is where Bing's foundation for safe and relevant search experiences begins. As part of product quality, Bing focuses on engineering and algorithmic design to ensure that users receive high-quality results that align with their search intent while limiting exposure to harmful or low-quality content.

The user's search query is the primary factor determining what information is displayed to users on Bing. The service prioritizes delivering content that is closely aligned with those queries, ensuring users receive information tailored to their individual searches. Unlike platforms that broadly recommend content based on user behavior across the service, Bing's approach centers on query-driven relevance.

When Bing surfaces suggestions, such as autocomplete, trending searches, or homepage modules; these are carefully filtered through targeted mitigations and quality thresholds. This ensures that safety standards are upheld and that harmful or inappropriate content is not elevated. This commitment to precision, relevance, and safety reflects Bing's dedication to maintaining high product quality and providing authoritative information aligned with user intent.

Within Product Quality, Bing's algorithms are the first and strongest defense in delivering safe, authoritative content. These systems are designed to match user queries with the most relevant third-party web pages in Bing's index, elevating trusted sources and reducing visibility of unreliable or manipulated content. Rather than removing content outright, Bing relies ***on continuous improvement of ranking algorithms*** to ensure that harmful content is not surfaced prominently, particularly when users have not explicitly searched for it.

To support this, Bing invests continuously in refining how its crawlers and algorithms detect quality content. These improvements are applied across all markets and languages where Bing operates. Teams regularly assess performance through key metrics, including internal safety indicators, and use these results to inform updates. In addition to technical signals, Bing incorporates feedback from users, Microsoft's policy teams, regulators, and civil society organizations to ensure its ranking decisions reflect evolving standards of trust and accuracy.

Bing's commitment to Trustworthy Search Principles means ***prioritizing high-authority content*** in search results, especially in areas where inaccurate or manipulated content may pose risks. Rather than removing content outright, Bing relies on ranking algorithms to ensure that harmful content is not surfaced prominently, particularly when users have not explicitly searched for it.

To support this, Bing uses a Quality and Credibility ("QC") score, which includes external signals that help identify authoritative sources. These signals help elevate trusted results while demoting pages that may be misleading or exploitative. Bing also uses ranking systems to defend against manipulation, including spam or search engine optimization abuse that could otherwise distort results.

These ranking systems are continuously measured and refined to ensure they reflect user expectations, policy goals, and emerging risks. The result is a more accurate and safer search experience that remains grounded in user intent.

Respecting ***user privacy choices*** is a core Microsoft commitment, and as part of Product Quality, Bing provides users with control over how their data is collected, used, and personalized. Through the Microsoft Privacy Dashboard, users can view, delete, or export their personal data, including search history. They can also adjust advertising and content personalization settings. For example, if a user frequently searches for travel content, Bing might personalize their Microsoft News feed with travel-related stories. These personalization features are optional and can be disabled. In the EU, Bing also requires users to opt in before using cookies or similar technologies that collect data for personalization.

Bing also offers users direct control over what they see through ***SafeSearch***, a feature that uses machine learning to automatically detect and filter adult and explicit content from search results. This feature is especially important in educational and household settings where additional safeguards may be required.

SafeSearch settings can be locked to prevent tampering and ensure consistent application of filters. Bing also provides a mechanism for users to report inappropriate content or suggestions. These reports feed into the continuous improvement of the underlying safety systems, helping Bing respond quickly to emerging issues.

Bing publishes and enforces its content standards through ***terms and conditions***, specifically through the Microsoft Services Agreement and the Bing Image Creator Terms of Use which, among other conditions, prohibit the generation or sharing of inappropriate or harmful content. These policies apply globally and are available in local languages. Bing reserves the right to suspend or ban users who attempt to misuse the service. For example, entering prompts designed to bypass safety systems is not permitted. These terms help ensure that Bing remains a safe and transparent service for all users.

Product Safety

The second layer of the Bing Safety Funnel is Product Safety; a focused set of engineering, policy, and testing mitigations designed to proactively identify and mitigate safety risks. Product Safety mitigations do not focus on content removal. Instead, they help users find high quality information for their query efficiently, for example by surfacing trusted information, adding context, or reducing the visibility of potentially harmful results. Bing's approach is not only deliberate, but also agile, ensuring it can respond quickly and effectively to emerging safety challenges across the service.

Engineering standards and processes establish robust technical foundations to assess and manage a variety of risks. These include rigorous testing protocols, expert reviews, launch readiness assessments, and in-product transparency tools that ensure safe and trustworthy user experiences.

Specifically, before any new feature or product is introduced to users, Bing requires it to pass through a robust ***pre-launch risk assessment***. This assessment process is part of Microsoft's broader compliance framework and includes reviews of privacy, security, accessibility, safety, and responsible AI concerns.

These reviews are conducted by professional compliance managers who ensure that products meet Microsoft's standards and legal obligations. Where risks are identified, teams must implement appropriate mitigations before the product or feature can go live. This process ensures that safety, privacy, and integrity are designed into Bing experiences from the start.

Bing also requires extensive ***red team testing*** to explore safety risks under pressure. These tests simulate how systems might be misused or manipulated in the real world. Multidisciplinary teams and, at times, invited external experts, rigorously test how Bing Search, Copilot Search, and Bing Image and Video Creator respond to adversarial inputs or edge cases.

Red team findings are used to refine training data, harden safeguards, and improve system behavior. Bing ensures that these tests continue even after launch, maintaining vigilance as systems evolve.

Every change to Bing, whether in core search or newer features like Copilot Search, must also pass a ***"shiproom" review***. This structured launch-readiness evaluation ensures proposed updates meet rigorous standards across safety, privacy, security, and accessibility.

Shiproom reviews require product teams to demonstrate the impact of proposed changes on Bing's core safety metrics, such as DDR. Teams must provide evidence of both offline and live (A/B) testing and explain how they've addressed any regressions or new risks uncovered during testing. Further, in the monthly shiprooms, issues and risks are presented to Bing leadership, enabling leadership ongoing governance and oversight. These checkpoints ensure that safety remains a central consideration in how Bing evolves.

Finally, Bing recognizes that safety includes not only what users are exposed to, but also how well they can interpret what they see. By providing contextual information, highlighting trusted sources, and promoting awareness of potential risks, Bing supports users in making informed decisions.

To this end, Microsoft promotes ***digital literacy*** across its services, including Bing. This work is essential in helping users critically assess online content, especially in the face of misinformation or potentially deceptive material, and complements Bing's other proactive safety efforts.

For certain high-risk query categories, Bing displays ***PSAs*** at the top of search results. PSAs highlight authoritative information including reporting and support options, and/or offer clear warnings about potential risks. There are, for example, PSAs in the EU related to the illegal nature of child sexual abuse material and resources and support for queries related to suicide & self-harm. PSAs are one of Bing's most visible tools for helping users navigate sensitive or dangerous topics with clarity and care.

These messages are localized, translated, and adapted to serve users in specific markets where Bing operates. Whether it's a PSA or a warning label, the goal is the same: to provide clear, immediate context when users encounter results that may carry risk.

Bing's safety philosophy includes keeping users informed when safety-related actions are taken. This includes ***in-product indicators and disclosures*** such as warnings, disclosure footers, or icons that explain when content has been removed or flagged, and why. For example, Bing may include a notification in search results when content has been removed due to legal or policy requirements. In AI-powered experiences, Bing includes clear disclaimers that the user is interacting with an AI system, along with reminders to verify content using authoritative sources. These disclosures are written in plain language and designed to be understandable even by younger or less technical users.

Together, these measures, and the full set of Product Safety mitigations in Appendix II, are thoughtful, precise, and continuously evolving. Consistent with the unique risk profile of a search engine, Bing's approach to product safety focuses on measured mitigations, transparency, and ongoing monitoring, rather than removing content or restricting access to information, to support a safe and empowering experience for every user.

Targeted Interventions

The third layer of Bing's Safety Funnel is Targeted Interventions. Targeted Interventions are a combination of algorithmic and manual intervention used by Bing when search results may not align with its Trustworthy Search Principles. These interventions are carefully designed and precisely deployed to address intentional attempts to manipulate search results or to respond to emerging threats in a way that protects users while preserving access to information.

As a search engine, Bing intervenes in limited ways, consistent with Bing's Trustworthy Search Principles to help users find relevant, credible, and safe information while mitigating the potential of unexpectedly encountering results that are harmful, offensive, or illegal.

As part of the ***Safety Measures – Defensive Search Interventions*** mitigation, Bing has a dedicated team accountable for implementing algorithmic defensive search interventions (as well as metrics monitoring and remediation) to address high impact issues in search results, such as information manipulation, hate speech, and other problematic content that could negatively impact user safety.

Bing uses query classifiers to identify queries that may be associated with elevated risk, such as those that could surface degrading, harmful, or exploitative content, and those trigger a set of defensive search interventions. These algorithmic measures are integrated rather than standalone and work together to uphold Bing's relevance, quality, and credibility standards—particularly in cases where harmful content may exploit data voids or manipulate user intent.

When a query is flagged as potentially risky, Bing may respond with measures such as boosting high-authority sources or applying algorithmic dampening to deprioritize low-quality or harmful content. These interventions are applied in accordance with Bing's core search principles and are actively monitored by the dedicated team that manages implementation and assesses impact across high-risk areas. This combined approach allows Bing to detect potentially harmful search patterns early and respond with targeted mitigations that preserve both user safety and access to high-quality information.

In cases where Bing detects threats to search quality in fast-moving or sensitive areas, Bing will ***direct users to authoritative sources.*** This includes areas such as elections, public health, or active manipulation of the information environment. Bing may demote low-authority content, boost more credible sources, restrict problematic autocomplete suggestions, or make manual adjustments to search results to limit harm. Each step is narrowly applied, with the goal of improving user safety while preserving open access to information.

To support this agile defensive work, Bing has a ***Threat Intelligence*** team that specializes in identifying and responding to high-risk search topics. This team continuously investigates where and how search results might be falling short and applies corrections with precision. The team toolkit includes boosting the visibility of authoritative websites, demoting those with low reliability, and restricting suggestions that could lead users toward harmful or misleading content.

Bing treats safety as a living system; constantly monitored and adjusted in response to new signals. There is ***ongoing monitoring of risks*** and metrics to assess whether interventions are working as intended. Bing uses a combination of internal red teaming, user reports, social listening, and threat intelligence to detect gaps and refine its approach.

These mechanisms allow Bing to catch new trends in attempted manipulation, identify emerging high-risk queries, and improve how its systems adapt to changing threats. If a mitigation unintentionally introduces bias or underperforms in a particular region, the team takes swift action to correct it.

Bing maintains ***social listening*** pipelines where insights and user feedback on Bing's features are collected from the Internet. These insights and user feedback are manually reviewed, analyzed daily, and shared across Bing's product teams and product engineering leadership in a daily report. These reports inform Bing as to the public concerns, issues, or emerging trends and can serve as barometer of public sentiment on various topics related to Bing.

Bing's Targeted Interventions helps mitigate the risk of users expectedly encountering harmful, offensive, or illegal content. The goal is not to censor or overcorrect, but to intervene in targeted ways leveraging a range of resources to support users' freedom to explore, question, and discover.

Content Moderation

The next layer in the Bing Safety Funnel is Content Moderation, which consists of narrowly scoped algorithmic and manual interventions applied when search results conflict with Bing's Trustworthy Search Principles due to legal, regulatory, or policy-based reasons. This layer is used with restraint and precision, focusing only on legal obligation and the clearest and most sensitive areas of harm.

In certain jurisdictions, Bing is legally required to remove access to specific search results. These laws may permit individuals, organizations, or governments to request the removal of indexed pages for reasons such as copyright infringement, libel, defamation, unauthorized publication of personal information, or other content that may violate personal rights or public order. Where such requests are submitted, Microsoft carefully evaluates each one to determine its legal basis, the scope of the request, and the requesting party's authority.

Bing's legal review process aims to strike the right balance between compliance with local laws and the preservation of fundamental rights to freedom of expression and access to information. Legal removal areas where Bing applies a consistent global approach include the removal of child sexual exploitation and abuse imagery ("CSEAI") and enforcement of copyright takedown requests.

Bing maintains rigorous and principled processes, policies, and procedures for ***content removal and other mechanisms to counter harms***. Central to these processes is a firm commitment to the prevention of CSEAI. Bing proactively scans content to detect known CSEAI using hash-matching technologies such as PhotoDNA and MD5. This scanning occurs before content enters the search index and is used to ensure such material does not appear in results at all, across Core Search and Copilot Search. Bing also proactively scans image uploads in Bing Image Creator and Bing Video Creator to prevent known CSEAI from being uploaded. This proactive detection is supplemented by intelligence from trusted partners, including NGOs and law enforcement, and through user reporting channels. When CSEAI is identified, Bing removes it globally and without hesitation. This is a zero-tolerance area where Microsoft's commitment is absolute.

Bing also works to remove adult or sexually explicit images of another person online when shared without that person's consent. This is commonly referred to as NCII, or "revenge porn". Microsoft considers this to be a gross invasion of personal privacy – whether the content is real or a "deepfake" image that is created using AI or other photo editing tools

Finally, Bing also removes certain other content under global legal standards or Microsoft policies, such as materials containing extremely sensitive personal information that could be exploited for fraud or identity theft. Removals may be prompted by legal orders from governments, user-generated reports, or policy-based decisions following internal investigations. Microsoft applies rigorous standards to each request, including the potential impact on user rights and public interest.

Bing provides users with an accessible ***report a concern*** mechanism on every page of Bing Search and across GenAI experiences. Through this function, users can raise concerns about potentially unlawful or harmful content. These reports are routed to human moderation teams who assess submissions based on Microsoft's policies, relevant laws, and the specific context of the reported material. While not all reports lead to content removal—especially in cases involving lawful but offensive or disputed content—each report contributes to Bing's broader understanding of emerging risks and informs future updates to its safety interventions.

Bing's commitment to privacy and dignity is also reflected in its implementation of the ***Right to Be Forgotten ("RTBF")***. Under this policy, users in jurisdictions such as the EU can request the removal of search results associated with their name if those results are outdated, inaccurate, irrelevant, or excessive. Requests are

reviewed by trained staff using clear criteria to balance the individual's right to privacy with the public's right to access information. Where appropriate, Bing removes URLs and associated search suggestions from the index in response to validated requests. Escalation pathways are in place for complex cases, including collaboration with local courts and data protection authorities where necessary.

This process is underpinned by strong training protocols and quality assurance systems. Bing ensures consistency and fairness in RTBF decisions, while maintaining transparency about the process through user guidance and regular engagement with regulatory authorities.

Bing's content moderation policies are shaped by Microsoft's broader commitments to digital safety, freedom of expression, human rights, and compliance with international law. While moderation is applied in limited and clearly defined scenarios, these efforts are critical to protecting users and maintaining the trustworthiness of the search experience.

Incident Response

Incident Response forms the final and most dynamic layer in the Bing Safety Funnel. It is designed to respond swiftly and effectively when urgent quality or safety concerns arise. These manual interventions come into play when internal monitoring, external escalation channels, or user reports identify issues that require immediate attention to uphold Bing's Trustworthy Search Principles.

Unlike earlier layers that rely on ranking systems, safety engineering, or defensive interventions, to guide safe and high-quality experiences, Incident Response serves as Bing's rapid-response capability. It reflects Bing's commitment to thoughtful stewardship of search results; which ensures users can trust that when issues emerge, they are addressed with care, speed, and accountability.

At the core of ***Incident Response and Service Level Agreements (SLAs)*** is a well-established infrastructure for responding to notices involving illegal or policy-violating content. This includes materials such as CSEAI, NCII, exposed personal information, and other high-risk violations. When Bing receives a credible report from governments, users, or trusted third parties, it acts swiftly to evaluate and address the concern.

Bing's support teams are trained in legal and policy standards, with clear procedures for escalating complex cases to local legal experts and subject matter specialists. Reports are tracked through Microsoft's internal incident management system, which ensures accountability and drives collaboration across engineering, legal, safety, and policy teams. Alerts are sent to designated stakeholders, and all incidents are triaged according to priority, with time-sensitive issues handled under strict SLAs that form part of Bing's broader Objectives and Key Results (OKR).

The prioritization process considers the nature of the report, the sensitivity of the policy area, and contextual details such as region, language, and media format. While response times vary based on these factors, most reports involving high-priority concerns are resolved within 24 hours. This fast-paced yet thoughtful process ensures that Bing's response mechanisms remain both effective and user-respecting.

Regular testing and quality checks are in place to ensure that these workflows operate reliably. Bing's commitment to continuous improvement means that incident response systems are refined over time based on lessons learned, stakeholder feedback, and evolving safety challenges.

Because Bing is a search engine, not a content-hosting service, its users do not post content directly to the service. As a result, Bing does not maintain user-facing ***appeals*** processes for content removals. However, where Bing applies enforcement actions, such as when removing or downranking search results in accordance with policy or legal obligations, it offers webmasters a means of appeal.

Through Bing Webmaster Tools, site owners who believe their content has been incorrectly impacted can request a review. These appeals are evaluated by trained reviewers, supported by escalation paths to experts where needed, to ensure outcomes are consistent with Bing's principles and commitments to fairness, transparency, and trust.

In addition to its core search product, Bing also supports services such as advertising that involve direct user participation. In these contexts, Bing enforces clear standards to prevent misuse that could harm users or compromise the integrity of the service. Advertisers who engage in fraudulent, misleading, or illegal behavior are subject to immediate ***suspension of advertiser accounts***.

Bing applies these actions judiciously and in alignment with Microsoft's broader safety commitments. Suspension decisions are made based on evidence and internal enforcement guidelines, with appropriate documentation and oversight to ensure consistency and due diligence.

Cross-Funnel Research, Partnership, Transparency, and Reporting

In addition to the five layered system of mitigations, Bing engages in research and partnership with a variety of internal and external experts to inform its standards, policies, practices, and processes. Bing is also transparent about its practices and assesses and reports on safety through various voluntary and regulatory channels. This is illustrated through the mitigation measures described herein.

Research and Partnership

Bing collaborates closely with ***dedicated cross-functional teams of experts*** across Microsoft. These partners specialize in areas such as digital safety, election integrity, and RAI. They help develop and maintain company-wide policies that guide Bing's decision-making on safety interventions and mitigation strategies. Their input ensures that Bing is aligned with evolving regulatory expectations and public interest standards.

Bing and Microsoft, including through its research arm ***Microsoft Research***, have collaborated with and third-party research organizations to contribute to novel research and internal studies concerning safe design practices, responsible AI, and information manipulation. Bing is evaluating additional research data sharing opportunities pertaining to elections.

Both Bing and specialized cross-Microsoft teams regularly ***engage externally with stakeholders*** in areas of key policy priorities, such as violative content, information integrity, responsible AI and AI-specific risks, minors, and technology, to help ensure that Bing internal policies, practices, and standards are addressing key concerns of these stakeholders. These engagements can inform the processes of identifying, assessing, and mitigating risks across Bing and provide greater understanding of concerns related to the Bing service and broader societal and technology related issues. External engagement gives Bing opportunities to hear feedback from a range of civil society, non-profit, researchers, and government stakeholders and learn from others in the industry.

Bing operates a qualified ***researcher program to enable EU researchers*** to easily request access for publicly accessible Bing data from a singular landing page. In addition, Bing is continuing to evaluate additional data sources and tooling to support the research community and looks forward to supporting the DSA vetted research program consistent with the recent Delegated Regulation.

Bing regularly ***engages with regulators*** around the world to understand key concerns, share information, and incorporate feedback into product design and safety systems as appropriate.

Furthermore, Microsoft has developed tools and multistakeholder ***third party partnerships*** to combat the rise of misinformation that can pose risks to public health or negatively impact physical or mental wellbeing.

Partnering with independent third-party organizations empowers Bing product teams to take additional actions to promote more authoritative information. Microsoft is also a founding member of the Coalition for Content Provenance and Authenticity ("C2PA") alongside Adobe, Intel, TruePic, Twitter, the BBC, and other tech and media companies. The coalition has an open-source content provenance tool, which allows creators to claim authorship while empowering consumers to make informed decisions about what digital media to trust.

As one example, Microsoft worked with independent third-party partners to perform red team testing of Bing Image and Video Creator to understand the risk of misleading images generated by Bing Image and Video Creator. As part of its analysis, the independent third party prompted Bing Image and Video Creator to create visuals that reinforced or portrayed prominent false narratives related to politics, international affairs, and elections. Based on these evaluations, Microsoft improved existing Bing Image and Video Creator mitigations.

Microsoft, including key members of the Bing Search team, actively participates in the ***Partnership on AI ("PAI")***, collaborating with peers across industry and civil society to identify and mitigate potential harms associated with synthetic media. As part of this work, Microsoft has contributed to the drafting of the PAI's Synthetic Media Code of Conduct; which are guidelines that promote the ethical and responsible creation, distribution, and use of AI-generated content, including deepfakes and digitally altered media.

Transparency and Reporting

Bing practices transparency by notifying users through notices on the search engine results page when content is removed and by publishing ***transparency reports***. Microsoft's Reports Hub is a publicly available source where users can access various transparency reports in areas where Bing reports on various practices. These key areas include RAI, content removal requests (such as copyright or digital safety), privacy (such as "Right to be Forgotten") and security (such as Government Requests). The Reports Hub contains additional transparency reports such as jurisdictional, community and privacy and security transparency reports that users can easily access. As a result of the conduct of the Systemic Risk Assessment, each year Bing identifies specific areas for focused enhancements in the coming year.

Bing provides ***user transparency*** and more detailed information on the main parameters it uses for ranking in the How Bing Ranks Your Content section of the Webmaster Guidelines, specifically calling out that in measuring the "quality" of a website, "pages that call for violence, name-calling, offensive statements, or use derogatory language to make a point are generally considered low quality." Bing works to ensure these resources are available in relevant languages and markets.

Users can learn more about Microsoft's collection and use of personal data in the ***Microsoft Privacy Statement***. The Microsoft Privacy Statement outlines what personal data Microsoft collects, how it is used, purposes for which it is used, and how to access and control personal data. This includes how these components apply to specific Microsoft products like search and browse, cookies, Microsoft Account information, and minors' data. Additional topics about personal data are covered such as security, storage, and retention, and how to contact Microsoft.

The ***Microsoft Bing Ad Library*** is another initiative aimed at enhancing user understanding and control over the advertising experience. This resource allows users to delve into the specifics of the ads displayed on Bing, including details about the advertisers, their expenditure, and the rationale behind the presentation of certain ads. The "Why This Ad" feature further demystifies the ad selection process by explaining the relevance of ads based on individual online behavior and interests, thereby offering users greater insight and authority over their digital environment.

Bing is a signatory to the ***EU Code of Conduct on Disinformation ("COCD"), formerly known as the Code of Practice on Disinformation,*** and subscribes to many commitments in the code dedicated to reducing disinformation risks, including providing users with indicators of content provenance, supporting good faith research into disinformation by enabling researcher access to data, and providing biannual reporting pursuant to the code on systems and policies in place to prevent disinformation on Bing across the EU, among others. Bing also provides enhanced reporting on agreed upon elections and other declared "crisis" under COPD.

As a search engine, Bing is committed to supporting and ***prioritizing users' fundamental rights*** and their rights to access and seek information. Bing does not broadly remove content from its index. Instead, it focuses on minimizing unintended exposure through ranking and targeted filtering.

Generative AI Experiences Specific Mitigations

Given the unique and evolving nature of GenAI capabilities GenAI receives dedicated consideration within Bing's risk management framework. Like other transformational technologies, harnessing the benefits of AI is not risk-free. A core part of Microsoft's RAI program and Standard is the requirement to identify potential risks, measure their propensity to occur, and build mitigations to address them. These risks are addressed through a defined set of mitigations, including:

Product Quality

GenAI Terms and Conditions governing the use of Bing's GenAI features are outlined in the Bing Image Creator Terms of Use. These provisions prohibit activities that could negatively impact users, such as engaging in harmful, fraudulent, or illegal activities, violating privacy, or creating inappropriate content. Users who violate these terms may face limitations on their access to the services.

All Microsoft products that incorporate artificial intelligence, including Copilot Search and Bing Image and Video Creator, are governed by ***Microsoft's RAI Standard***. This standard requires comprehensive reviews and mitigation planning before AI-powered features are launched.

These reviews focus on identifying potential risks, assessing the likelihood and severity of harm, and ensuring mitigations are robust, context-appropriate, and in line with Microsoft's AI Principles. For GenAI experiences, this often includes added layers of technical safeguards, transparency mechanisms, and human oversight. By rigorously evaluating generative features before deployment, Bing prioritizes user safety and dignity, while still enabling innovation. This process reflects Bing's commitment to maintaining a respectful, secure digital environment in the face of emerging AI capabilities.

To guide GenAI models in generating responses that align with Microsoft's AI Principles and user expectations, Bing employs a technique known as "***metaprompting***." Metaprompting involves providing instructions to the AI model to influence its behavior, ensuring that the system operates responsibly and ethically.

Copilot Search responses are generally ***grounded in high-ranking search results***, a process known as "grounding." This approach involves providing data to a large language model (LLM) to improve its ability to generate accurate, relevant, and updated outputs. By centering responses on high-authority content from the web, Bing aims to prevent the generation of content containing private information hosted on low-authority sites. Additionally, providing links to source websites allows users to evaluate the credibility of the information presented.

User-centered design and user experiences are essential aspects of Microsoft's approach to responsible AI. The goal is to root product design in the needs and expectations of users. As users interact with Copilot Search and Bing Image and Video Creator for the first time, these services offer various touchpoints to help them understand the capabilities of the system, disclose that they are powered by AI, and communicate limitations. Elements of the experience also help users better understand its interactions, including chat suggestions specific to RAI, explanations of limitations, ways to learn more about how the system works, and easily navigable references to show users the results and pages in which responses are grounded.

Copilot Search and Bing Image and Video Creator include ***GenAI in-service disclosures*** and reminders to users that AI can make mistakes, encouraging them to verify the information generated. Furthermore, Bing provides FAQs, help pages, and other public-facing information sources to educate users on the nature of AI-driven search experiences, including their uses, safeguards, and limitations.

Copilot Search responses include ***citation links***, and the corresponding web results are displayed at the bottom of the results. This practice allows users to verify the information and assess the credibility of the sources.

To address issues related to ***conversational drift***, Bing has limited the number of turns per chat session in Copilot Search. This limitation helps prevent responses that are repetitive, unhelpful, or inconsistent with the intended tone. The team continues to evaluate additional approaches to mitigate this issue.

In cases where a user's prompt is ambiguous, Copilot Search and Bing Image and Video Creator may use the LLM to build out more details in the prompt to help ensure users receive the response they are seeking. This ***prompt enrichment*** does not rely on any knowledge of the user or their prior searches but instead on the AI model. These revised queries are visible in the user's chat history and can be deleted using in-product controls.

Bing empowers users to make informed decisions about their ***privacy choices*** and how their data is collected and used in GenAI experiences. Through the Microsoft Privacy Dashboard, authenticated users can view, export, and delete their stored conversation history, including any queries or prompts submitted in Copilot Search.

These controls complement existing privacy settings, such as cookie banners and consent mechanisms, which continue to apply to generative features. Bing also applies an internal review process to examine how choices are presented, ensuring users clearly understand its options and the implications of its decisions. By giving users meaningful control over its data, Bing supports autonomy and accountability in the use of AI-driven technologies.

Product Safety

Transparency is essential in building trust in AI. For ***AI Transparency***, Bing displays prominent disclosures and in-product notices when users are engaging with AI-generated content. For example, in Copilot Search, an information icon near the feedback buttons alerts users that the response was generated by AI. Hovering over this icon reveals a plain-language explanation, with links to more detailed documentation. These disclosures are written at a fifth-grade reading level to ensure they are accessible to young users and audiences of varying digital fluency. To support informed engagement and reduce manipulation, Bing has also introduced content integrity as a Service, a system that uses digital credentials and cryptography to authenticate the origin of media. This helps users trust that the images or media they are seeing have not been tampered with, and that they come from verified sources.

Bing has certain ***age restrictions for GenAI features***. While some, limited functionality in Bing Image and Video Creator is available to all unauthenticated users, full functionality is restricted to authenticated users who meet the appropriate age thresholds. Users must be signed in with a Microsoft account that reflects an age of 13 or older, or a higher age in jurisdictions where parental consent laws apply, to access full functionality, including multiple turns on the same content. If a user's account indicates they are below the required age, they are prevented from accessing Bing Image and Video Creator, even the limited experiences available to unauthenticated users. This age-based access control helps protect younger users from exposure to content or interactions that may not be appropriate for development.

Bing Image and Video Creator uses targeted safeguards to limit the generation of inappropriate or potentially harmful content. A key intervention is the use of intelligent ***blocklists***—filters designed to detect and restrict specific user inputs that contain sensitive or risky names, terms, or phrases. These inputs are sorted into "hard" or "soft" blocklists, depending on the level of sensitivity and risk associated with them.

The development and refinement of these blocklists is a collaborative process involving internal product and safety teams, insights from jailbreak researchers, red teams, and specialized legal counsel. For example, during election periods, Bing consults with elections counsel to ensure the blocklists reflect heightened sensitivities around electoral misinformation or manipulation. This multi-disciplinary approach allows Bing to tailor safeguards to real-world risks in a dynamic and proactive manner.

The development and deployment of GenAI features at Bing are guided by safety processes that ensure adherence to high standards of responsibility and ***emphasize safety by design for GenAI features***. These processes include red team testing, expanded reactive social listening systems, prominent reporting functionality, and operations and incident response mechanisms. These measures help identify vulnerabilities and inform the development of targeted mitigations.

To address risks associated with Bing generating harmful content, Bing has placed specific, ***targeted limitations on GenAI text and image responses*** for certain queries. For example, when Bing classifiers identify potentially harmful content in user prompts for Copilot Search and Bing Image and Video Creator, Bing may not return generated content to users, may divert users to different topics, or redirect users to traditional search results. These measures help prevent the dissemination of harmful or inappropriate content.

Targeted Interventions

Bing deploys specialized mitigations in its GenAI features, such as "***classifiers***," to reduce the risk that harmful content might be generated by Bing. These classifiers analyze text, images, and videos to identify potentially harmful content in search queries and generated responses. By applying AI-based classifiers and content filters, Bing can flag and prevent the production of harmful content, including in features such as Copilot Search and Bing Image and Video Creator. For instance, classifiers have been implemented to mitigate the risks of generated content related to public figures and to prevent the generation of low-authority content.

Content Moderation

Copilot Search uses the same content moderation processes as Core Search, and users are able to report concerns through the same features and tooling. Additionally, Bing allows individuals, such as artists, creators, or brand owners, to request that Bing include their name on the ***Bing Image and Video Creator blocklist*** so that the GenAI models do not replicate their artwork.

Incident Response

In serious cases, Bing may ***restrict a user's access to generative services*** or suspend a user's account when there is evidence of repeated or intentional misuse. This includes attempts to bypass content safety protections or efforts to generate highly sensitive material. User's who demonstrate clear intent to create CSEAI will have their accounts suspended.

For Bing Image and Video Creator, users whose prompts are repeatedly flagged for policy violations may be automatically suspended for a defined period or permanently banned in severe cases. To support fairness and transparency, these users are given an opportunity to appeal such decisions.

Appeals are reviewed by trained human moderators who consider the user's behavior, applicable terms, and context. Based on the outcome, the enforcement decision may be upheld, or the user's access reinstated. This process ensures that even enforcement actions within GenAI experiences are guided by a commitment to procedural fairness and proportionality.

How Bing monitors effectiveness of risk mitigation efforts

As noted above, Bing is committed to data-driven governance and relies on measurements across the Safety Funnel to track and manage online safety risks across its services. Metrics are considered holistically across Bing's mitigations, such as the overall Defensive Defect Rate ("DDR"), Bing's core safety metric and the DDR for each DSA systemic risk area, as well as threat intelligence, defensive search interventions, user reports, social listening, that work together to evaluate the impact of mitigations, surface issues, and guide continuous improvement. Each plays a distinct role in the system.

Measurement metrics

Bing uses DDR as a primary means to measure the service of how often harmful or low-quality content appears in search results by using "adversarial" metrics sets that are designed to mostly contain high-risk queries that are prone to lead to harmful results to measure associated defects. Bing monitors its overall DDR, as well as DDR based on the DSA systemic harm areas. Bing sets objectives to improve overall DDR on an annual basis and reviews them on a monthly basis to drive continuous improvement. Once Bing has achieved the set objective, the bar is raised with a new adversarial set or stricter evaluation criteria. This culture of continuous enhancements has allowed Bing to keep improving the safety of the service over several years.

In addition to quantitative signal of safety performance of DDR, Bing's commitment to metrics spans its layered system of mitigations, and each has distinct value. For example, metrics related to red teaming and threat intelligence proactively identify new manipulation tactics; defensive search interventions capture the performance of targeted algorithmic responses in high-risk contexts; user reports offer direct feedback from real-world users; social listening captures broader sentiment and trends; and incident reporting provides visibility into the effective of Bing's response processes. This integrated approach, continually monitored and adapted based on new signals, enables Bing to both identify gaps quickly and adapt to emerging threats and assess the longer-term effectiveness of safety mitigations to inform future investment.

Bing Overview and Risk Profile

User Base and Market

Bing has approximately 133 million average Monthly Active Users ("MAU") in the EU during the period of August 2024 to April 2025. This figure reflects users located within the EU (as identified by IP address) who – at least once per month – query Bing and view the results page on a desktop computer.

Bing has both authenticated and unauthenticated users. With the exception of optional image and video generation features that require authentication, Bing features, including other generative AI experiences such as summarizing articles or documents, translating, writing, rewriting in different tones or styles, are available to both authenticated and unauthenticated users.

Most of Bing's services are available to users across ages, subject to parental consent requirements. Authenticated users under the age of consent in their local region require parental consent to use the service.

Bing's average monthly active user base is not necessarily an accurate proxy for the reach or impact of Bing, as a sizable portion of Bing users are "light users" that query Bing a few days per month. Bing research suggests this can include users who may primarily use other search engines (or social media in lieu of a traditional search engine) but, who once or twice a month enter queries into a Bing interface, such as through the Windows Start menu or Microsoft Edge browser. At the same time, Microsoft recognizes the potential for content risks with this smaller population of users and Bing takes steps commensurate with its risk profile and its status as a designated VLOSE to address digital safety for users in the EU and beyond.

Bing offers over 200 market versions and is available in over 100 different languages, including in all EU Member States as of July 2025.

Considerations for Bing's Risk Profile

When using a search engine, users can have many valid reasons for wanting to seek out legal content that could be problematic or harmful if encountered in other contexts. As one of the few services that can connect individuals with as-yet-undiscovered content on the World Wide Web, Bing plays an important role in upholding the fundamental rights of free expression and access to information. In this context, one of the highest areas of risk is in ensuring that Bing's algorithmic systems can connect users with the content they are seeking, while ensuring that its content policies and practices regarding this third-party content are sufficiently robust such that users are not inadvertently misled or harmed by search results.

Bing also collects data from its users, including search terms and related data like location and language preferences that help provide more useful search results, and recognizes its obligation to comply with applicable data protection laws and Microsoft policies to ensure appropriate stewardship of that data and minimize risks of harms related to misuse of personal data.

However, given that Bing generally does not host user content, allow for messaging between users, or allow users to publish content on the service, data-related practices such as communications, Bing tends to be a lower vector for systemic harms than may be present on other services. Because Bing generally does not allow users to post or share their own content on the service, unlike in social media, there is limited risk related to user behavior toward other users on the service, and limited risk on the service that sharing of user-generated content (or improper enforcement of rules related to user generated content) give rise to significant systemic harms.

Bing does not generally prohibit users from entering search terms that indicate an intent to find harmful, offensive, or potentially illegal content as users can have valid research reasons for seeking out various types of content in search, even content that could be harmful or illegal in other contexts. This means that while Bing does have established terms and conditions that apply to its users, and does enforce those terms where needed – such as in the case of a user attempting to use illegal CSEAI as a search prompt in visual search, or users who violate terms or attempt to jailbreak GenAI features – user behavior on the service is not a significant vector for harms on search, and enforcement of Bing's terms and conditions plays a relatively minor role in mitigating systemic harms in Bing.

The GenAI experiences in Bing were designed to provide new ways for users to find the information they want, making it easier for users to find answers to their research questions, and answer more sophisticated questions or inspire new creativity with images. Many of the risks are similar to traditional core search: ensuring the algorithms and AI systems underlying the service are designed to return the content users are seeking, enhanced by content moderation policies and systems designed to ensure users are not harmed by content they seek in search results. Users still cannot post or share content on these services, but Microsoft recognizes that there is an increased risk that users could take content generated using these tools and share it separately on third party services and platforms in harmful ways. As a result, Microsoft has implemented specialized AI mitigations (such as metaprompts, filters, and classifiers), enhanced incident response and monitoring, and additional terms and Codes of Conduct (and related enforcement processes) with its GenAI tools to ensure they are used in appropriate ways, and that enforcement actions are fair.

As an ad-supported service, Bing recognizes that there are risks associated with advertisements on the service and works to ensure that the advertisements appearing on the service meet Bing's standards for content delivery, in terms of its content policies and enforcement, its ad ranking practices, and data use as well as the standards of Microsoft Advertising.

Core Search Features

Core search refers to the "core" Bing search engine experience wherein users can enter queries to search the Bing index for relevant web results. Bing crawls the web to build an index of pages (or URLs) to display as a set of search results relevant to a user-initiated search or action. Complex algorithms generate Bing Search results by matching the user's search query with third-party webpages in Bing's index. In some scenarios, Bing may suggest search topics for a user based on their search history or trending topics, such as via search suggestions, and homepage content on the Images and Videos tabs. Bing designs its ranking and recommendation algorithms to align with core product principles that prioritize high quality, relevant content, and to reduce the risk of users encountering harmful or misleading content (absent explicit intent to look for such content). Bing also removes illegal content such as CSEAI when identified.

Given the vast number of websites on the Internet and the trillions of websites in the Bing Search index, the content of these pages may vary greatly and can include images, videos, audio, text, links, downloadable documents, or other materials. Bing does not host the content on the websites that appear in search results, but at most caches third-party pages to deliver search results more quickly to users. Bing has no control over the operation, design, or contents of the materials on third-party websites, and third-party websites are not content provided by "recipients of the service" within the meaning of the DSA Article 3 definition. As long as a website makes content available on the Internet and to search engine crawlers, the content will generally be available through Bing and other search engines. Bing does not allow users to post or share their own content on the service; rather, it allows users to find and consume information from a variety of third-party sources.

Core search may also provide users with additional features to help provide additional context and information and enhance the search experience. For example, if users want to know which team won a particular sports match, when users search on Bing for a team name, they could be presented with a special Answer including the final score and a recap of the most recent match at the top of the page, as well as an overview with more information about the team, including news, videos, and related searches along with the most relevant search results. To help users find what they are looking for more easily, Bing also lets users narrow their search results with categories like images, videos, news, and shopping. These features also rely on the ranking principles and main parameters of core search, but in some cases take additional considerations into account to provide users with the most relevant results.

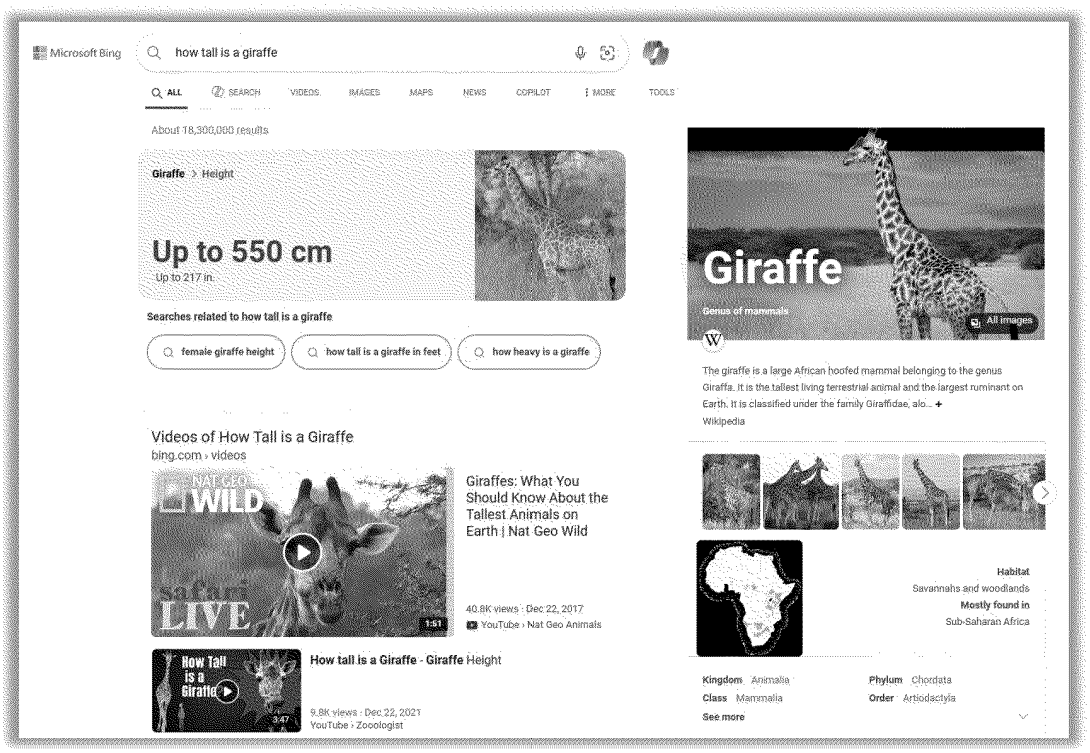
How Search Works

Returning search results involves complex, near-instantaneous algorithmic calculations. The first step in building search is figuring out which pages exist on the Internet to index them, which is a process commonly called crawling. Pages identified through Bing's crawler are added to the Bing index and algorithms are used to analyze the pages to effectively include them in search results, including determining which sites, news articles, images, or videos will be included in the index and available when users search for specific keywords.

Answers

"Answers" are enhanced search results provided at the top or side of standard search results that provide richer content in response to a user's query. For example, if a user types "How tall is a giraffe?" Bing will respond with the answer of "up to 550 cm. for males" to enable users to quickly find the information they need, along with a sidebar pane with more information about giraffes. Bing derives answers from high authority search results across the web and links to the original source. Bing may also offer specialized answers, such as in relation to elections or other high interest topics. Some of these answers may be powered by GenAI to provide an even richer experience, linking to key sources.

Figure 2: Core search results screenshot



Autosuggest and Related Suggestions

Autosuggest and Related Suggestions ("suggestions") are query suggestions provided by Bing to help users use search more conveniently. Suggestions is a feature that shows topics that the users might be interested in next to search results (e.g., "People may also ask"). Suggestions are generated and algorithmically ranked based on the popularity of related searches on Bing and natural language generation technology trained on query sets to help predict a user's intended query, along with other relevance signals such as search history, trends, location, and language.

Figure 3: Search suggestions screenshot

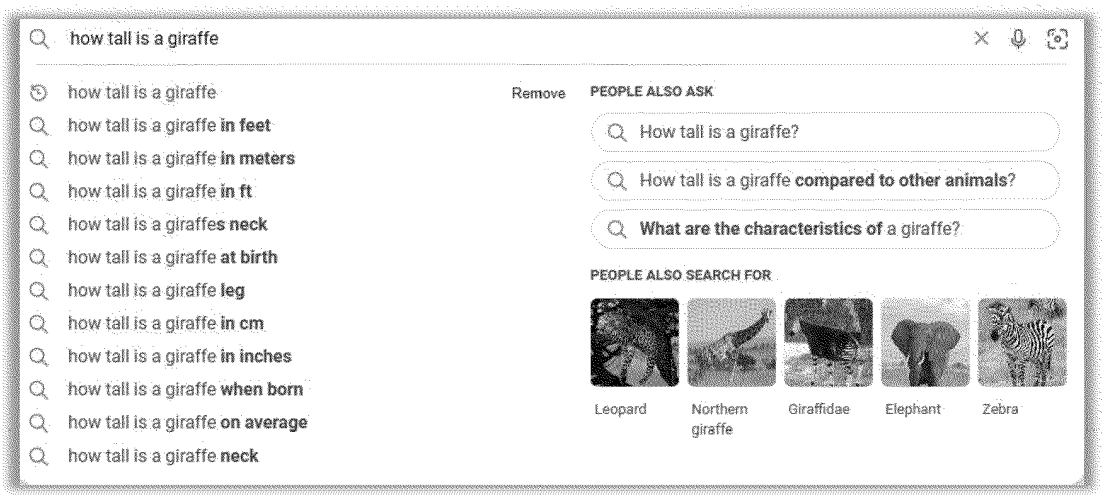
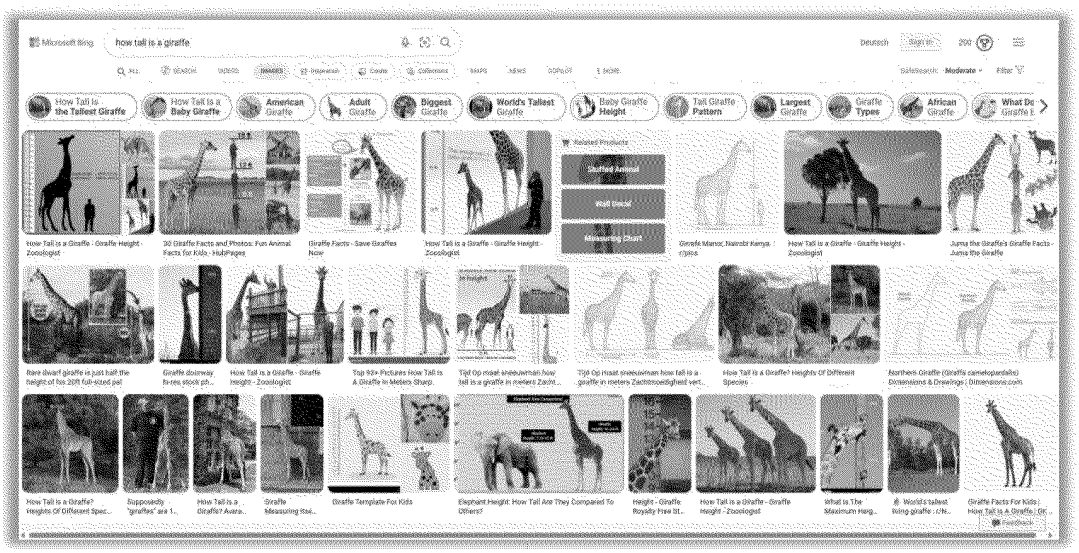


Image and Video

Bing's image and video experiences provide users with image and video results relevant to their search queries. These experiences can appear in the image and video verticals and on the main search results page. Ranking within these experiences generally relies on the same parameters as the main web search results page, e.g., relevance, quality, freshness, authority, and popularity.

Figure 4: Bing Image results screenshot

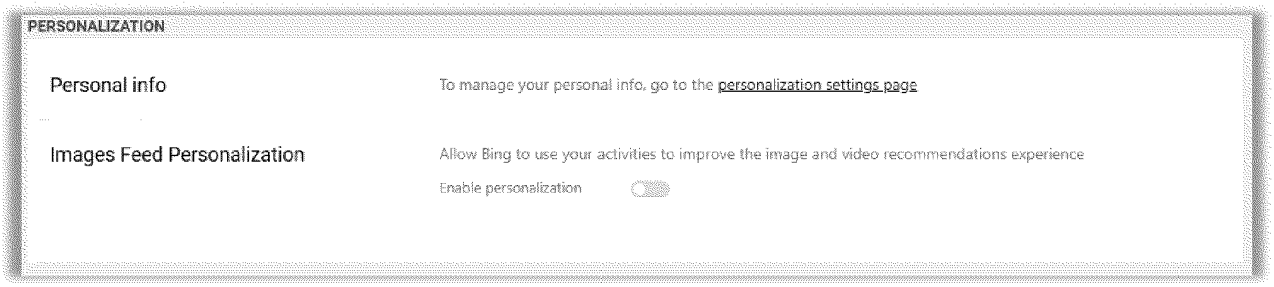


When a user first lands on the image or video experience homepages, prior to entering a query, the homepage may show recommended content based on prior interactions with the site, such as the user's search history, engagement with image results, saved images, and - where a user allows - their Edge browsing history. Users can control this experience by deleting data used to influence personalization from their Privacy Dashboard, leveraging the consent for Microsoft Edge browsing activity for personalized advertising and experiences, opting out of personalization in its entirety, or, if they are not logged in, starting a new session. Users can also toggle their image and video feed personalization through Bing's Settings.

Visual Search

In Visual Search in Bing, with an image taken by the user's camera, uploaded from the user's device, or linked from the web, users can prompt Bing to understand the context, interpret, and answer questions about the image. Users can also upload their files to Bing to interpret, convert, process, or calculate information from them.

Figure 5: Personalization toggle screenshot



MICROSOFT CONFIDENTIAL MSFT_JUD_000002467

Core Search Risk Profile

Search engines play a vital role in society considering the heavy reliance on search engines to access information that is critical for day-to-day life. In consideration of this role, search engines must carefully balance the fundamental right of freedom of information against digital safety. In some cases, a higher risk tolerance related to safety is warranted to ensure critical access to information.

Users approach search engines with questions and will typically see content that they have requested, rather than content that is shared or proactively recommended to them. Content on a search engine is not generally shared from user to user within the search ecosystem, which reduces the ability for content to "go viral" on the Bing service. Therefore, while referenced harms may be present within core search, the impact is unlikely to become systemic due to the limits on user-to-user sharing and recommendations.

Core search represents Bing's central function as a search engine and is the source of the vast majority of Bing's usership with the highest usage of any Bing feature or service.

Some ways that systemic risks can manifest on a search engine absent sufficient mitigations include:

- Risk that bad actors manipulate search results or exploit data voids to surface low authority, low quality content with an intent to deceive or manipulate audiences.
- Risk that search engines return content that provides hateful rhetoric or advocates the use of violence against people because of their race, or other protected traits in the top search results.
- Risk that search over moderates the search results, removing or downranking URLs limiting access to information.
- Risk that search engines return otherwise illegal content in the top search results.
- Risk that search engines return potentially distressing health or other information that can have a negative impact on users' mental health.
- Risk that search results or autosuggest features lead users to content that includes NCII or leads users, including minors, to synthetic NCII content.
- Risk that users are harmed by inappropriate processing of personal data, such as lack of transparency, control, access to remedies, or disclosures.
- Risk that search engines return content that is inaccurate, misleading information, harmful misinformation, related to crisis events or issues of public security in the top search results.
- Risk that search engines return content that instructs on, normalizes, promotes, or facilitates sexual exploitation, grooming, or abuse of a child.

The key mitigation measures Bing has implemented to address these potential risk manifestations are described in the Bing's Approach to User Safety section.

Generative AI Experiences

GenAI experiences offered via Bing include Copilot Search, Bing Image Creator, and Bing Video Creator, as well as a variety of embedded experiences such as summarizing articles, translating, writing, and rewriting content in different tones and styles.

These AI experiences are in part powered by Large Language Models (LLMs) developed by Microsoft's technology partner Open AI, including GPT, a cutting-edge LLM, and DALL-E, a deep learning model to generate digital images from natural language descriptions. Microsoft GenAI tools, including Copilot Search and Bing Image and Video Creator, are operated independently of OpenAI, and Microsoft has added

additional safety mitigations to the underlying models, including combining with search algorithms, adding bespoke metaprompts to guide GenAI responses, and additional safety filters and classifiers.

Responses are presented to users in several different formats, such as traditional links to web content, AI generated summarizations, and images. Summarizations that rely on web search results include references and a "Learn more" section below the responses, with links to search results that were used to ground the response. Users can click on these links to learn more about a topic and the information relied upon in the summary or chat response.

Copilot Search

Copilot Search seamlessly blends the best of traditional and generative search together to help users find what they need and represents a significant advancement in transforming the search experience. Copilot Search is an update to the previously shipped product, Real Time Generative SERP (search engine results page) in Deep Search. Copilot Search, like Real Time Generative SERP in Deep Search before it, responds to the user query with Generative Text. The text is generated by using an LLM to understand the intent of a user's query, issuing additional search queries to Bing to understand the original user query, and then synthesizing a response which is grounded on the retrieved search results. Real Time Generative SERP and Deep Search no longer exist as they evolved into Copilot Search.

With respect to features and functions, Copilot Search is similar to Real Time Generative SERP in Deep Search experience. Depending on the query, the user will receive an easy-to-digest summary of the most critical points, a clear answer, or a smart layout of information to help them find what they are looking for. Bing has set out to deliver helpful, clearly cited sources, plus rich, relevant data, images and videos from the users' favorite publishers and content owners to inspire the user to explore further. There is no modification to ranking of web sites in the search results that are used by grounding.

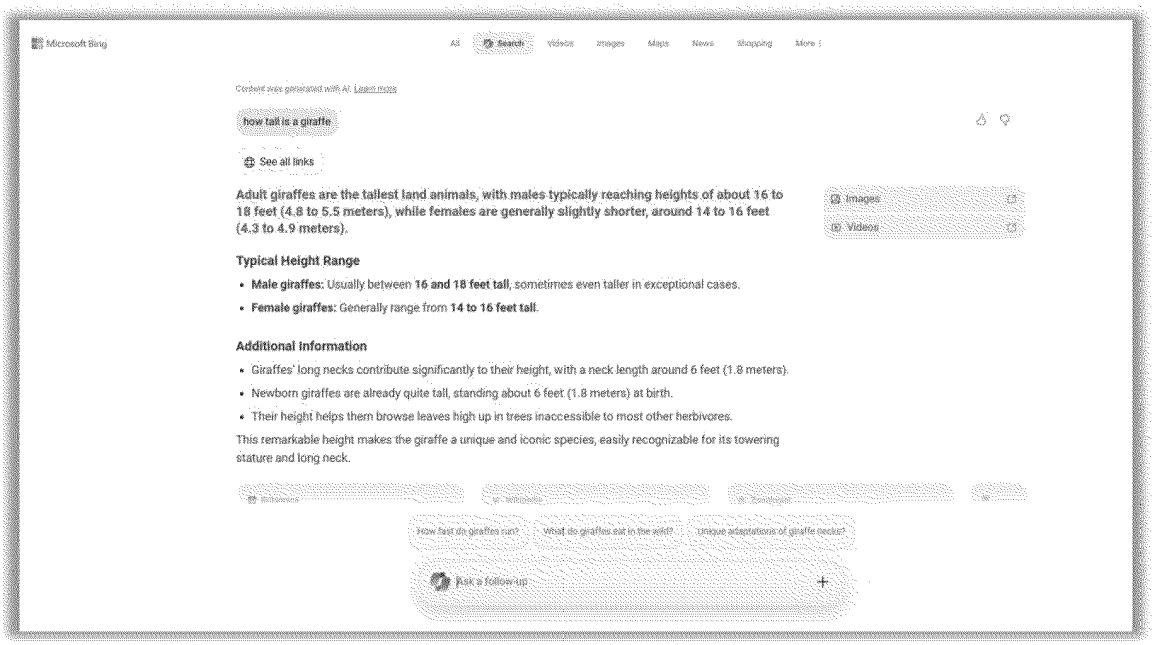
Bing has designed Copilot Search with publishers and content owners in mind to support a healthy web ecosystem. In addition to the citations within the generative responses, Bing inline links the entire sentence or passage within the responses so the user can easily navigate to the sources. Bing also includes easy access to cited sources and all relevant web results at the top and they're also highlighted prominently at the bottom of the result. This allows the user to be just a click away from their favorite publishers and content owners. Each Copilot Search result will have suggestions on related subjects. These clickable topics help the user uncover new information and make deeper exploration easier than ever – all without losing context in a multi-turn experience. Once a user clicks a topic, results are quickly summarized, and additional follow-up searches will be suggested so the user can continue exploring. Plus, in Copilot Search, the users' previous result is accessible from the same results page, making it simple to compile information around a single topic in context.

New features that are added to Copilot Search include the ability for the user to ask follow-up questions / queries about the generated response. These follow-up questions can be either free-form user queries or pre-defined suggestions. Users that are not signed in have a 30-follow-up question / query limit per day whereas users that are signed in have a 100-follow-up question / query limit per day. Also, in Copilot Search, depending on the query, there can be different experiences contained within a single page, including but not limited to a GenAI text summary with references, sections of creative text (creative writing, code generation, math), or a creative image, supported by Bing Image or Video Creator. This image creation functionality is the same as what exists on bing.com.

Copilot Search will in some instances, not provide a GenAI experience because responses may be inconsistent with Bing's policies and/or applicable law. In these cases, Copilot Search will return traditional core search results related to the user prompt.

Users can access Copilot Search by going to aka.ms/CopilotSearchinBing and start exploring. Users can also access Copilot Search via bing.com either through the navigation bar or simply by clicking on the suggested, related topics right under relevant answers in Bing.

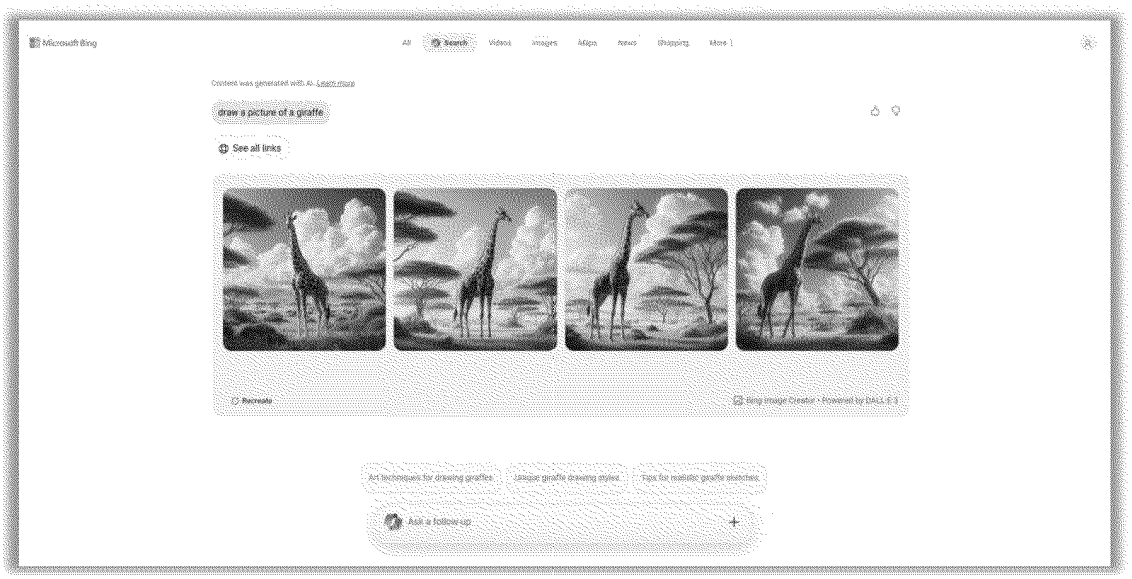
Figure 6: Copilot Search results screenshot



Bing Image Creator

When Bing Image Creator receives a user request to generate an image (for example, the prompt begins with "draw me a ..."), the system will, in most cases, generate an image response to the user's prompt. By typing in a description of an image and other context, such as artistic style or key image details desired, Bing Image Creator will generate a response, which is then displayed. For example, a user can type "create an image of a giraffe" as a prompt in Bing Image Creator to generate graphics depicting a giraffe. Bing Image Creator relies upon an image generation model from OpenAI, with additional Microsoft safety measures implemented. With Bing Image Creator, users can not only generate images using prompts but can also resize or restyle them by choosing from different restyle options. Authenticated users signed in with a personal Microsoft account have 15 "boosts" (fast creations) a day, after which image generation will move to a slow queue resulting in longer generation time. Authenticated users signed in with a work or school account are not able to create further images after they use all of their boosts.

Figure 7: Bing Image Creator results screenshot



Bing Video Creator

When Bing Video Creator receives a user request to generate a video the system will, in most cases, generate a visual video responsive to the user's prompt. By typing in a description of a video and other context, such as artistic style or key frame details desired, Bing Video Creator will generate a response, which is then displayed. For example, a user can type in "create a video of a giraffe walking" as a prompt for Bing Video Creator to generate a video. The generated video is a dynamic set of frames that does not have any audio.

Generative AI Experiences Risk Profile

While users can create AI-generated content using GenAI features available on Bing, and users can choose to download or share this content via separate means, this AI-generated content is not indexed by the core search feature nor disseminated through the Bing service. The nature of GenAI features and Bing's non-dissemination of the material is a substantial consideration impacting Bing's risk profile, as one user's generated content is not automatically surfaced to other GenAI or search users.

Some ways that systemic risks can manifest on GenAI features absent sufficient mitigations include:

- Risk that GenAI summaries are based on low authoritative sources and produce content that may be harmful or misleading.
- Risk that adversarial user queries or grounding data may bypass the safety features for GenAI and produce harmful or misleading content.
- Risk that GenAI Protection Systems flag and block the generation of legitimate content.
- Risk that GenAI summaries inaccurately interpret source data to generate incorrect or misleading information about an individual that could be defamatory.
- Risk that oversimplified GenAI summaries generate potentially distressing health or other information that can have a negative impact on users' mental health.
- Risk that GenAI returns content that could be used to find applications, tools, or services for creating synthetic NCII in the top search results.
- Risk that GenAI returns content to minors that normalizes eating disorders in the top search results.

- Risk that bad actors misuse Bing Image or Video Creator through generating images that are used to intimidate, harass, or coerce political or civic actors into complying with their demands.
- Risk that Bing Image or Video Creator is used to produce images or videos that degrade or discriminate against specific groups.
- Risk that Bing Image or Video Creator disproportionately generates images that align with particular cultural standards or viewpoints, resulting in less diverse representations of the population.
- Risk that Bing Image or Video Creator could be misused to generate synthetic, yet realistic private or intimate images of individuals.
- Risk that Bing Image or Video Creator is used to generate images that contribute to public security risks.

The key mitigation measures Bing has implemented to address these potential risk manifestations are described in the Generative AI Experiences Specific Mitigations section.

Ancillary Search Features

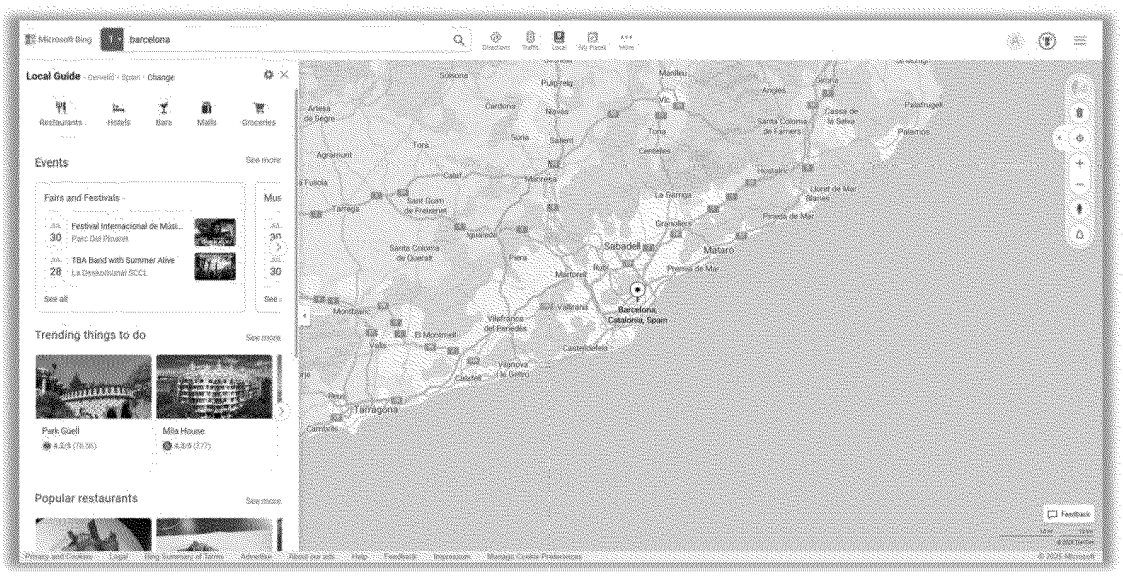
Bing Core Search links to a number of "verticals" or features areas to enable enhanced search functionality, including:

Maps

Bing Maps allows users to search for information about businesses, landmarks, locations, and other geographic data in a map interface. In France and Germany, Bing Maps offers a Local Guide feature where users can view hotels, bars, things to do, restaurants, and events. Businesses can improve the accuracy of Bing Maps results by registering with Bing Places and providing up-to-date information for Bing users on information relevant to that business, such as location, contact information, and hours. Users can also suggest locations to be added to Maps, which then undergo automated and potentially manual review before being made visible to other users. Users can post photos to businesses, which also undergo automated reviews for harmful content.

Maps' Commute feature allows users to identify their home and work to provide a personalized route. With the destinations feature, users can add other destinations to their personalized routes, for example a school or gym, to get directions quickly between their home, work, or other top destinations. Within the Commute feature, users can receive personalized traffic news. The user's location is sent without an identifier and used to fetch relevant traffic news. The location is not stored on a server. Users can also choose to share their Estimated Time of Arrival using various 1st and 3rd party apps on their mobile device. Personal data is not shared with these apps unless the user explicitly chooses to share it.

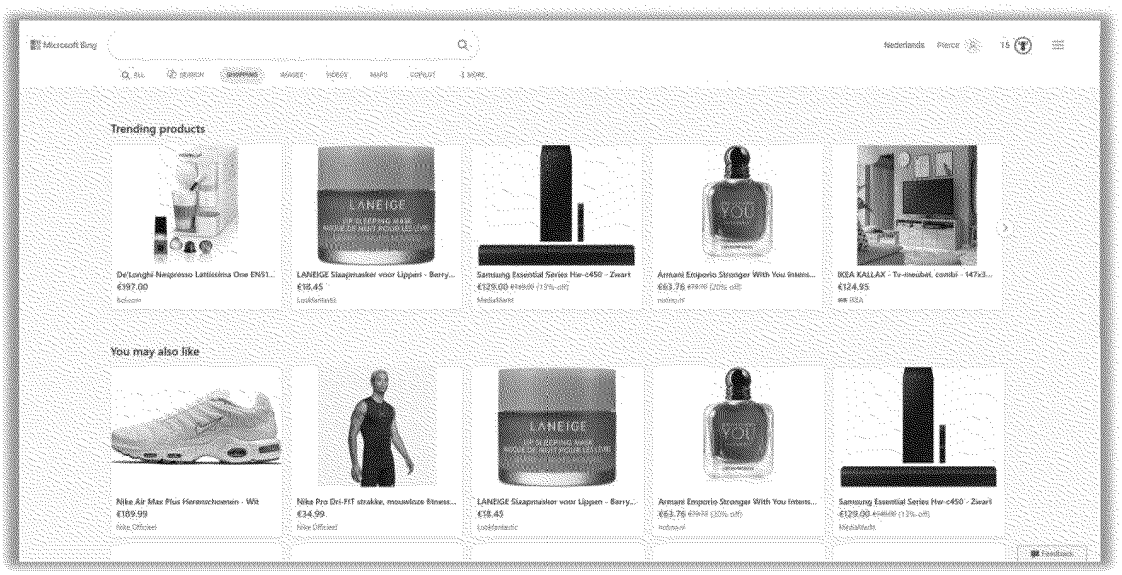
Figure 8: Bing Maps interface screenshot



Shopping

Bing Shopping is a search feature to help users discover products available through third-party websites by returning relevant shopping results from the web as well as sponsored advertisements relevant to a user's query. As with web results, users in the EU must click through to the third-party websites offering products for sale to complete a purchase. Users do not make purchases directly on or from the Bing service, and Bing does not process payment. In some cases, the shopping experience may use GenAI tools to provide better results for users. In some instances, Microsoft may receive compensation for users who click on algorithmically generated results, but such compensation has no effect on the ranking or relevance of algorithmically generated results shown to the user. Users who allow Microsoft to collect and use personal data to personalize advertisements may see results in the Shopping vertical tailored to their interests. Users can visit the Microsoft Privacy Dashboard or the Edge shopping toggle to opt-in or out of personalized advertisements and control the data that influences personalization.

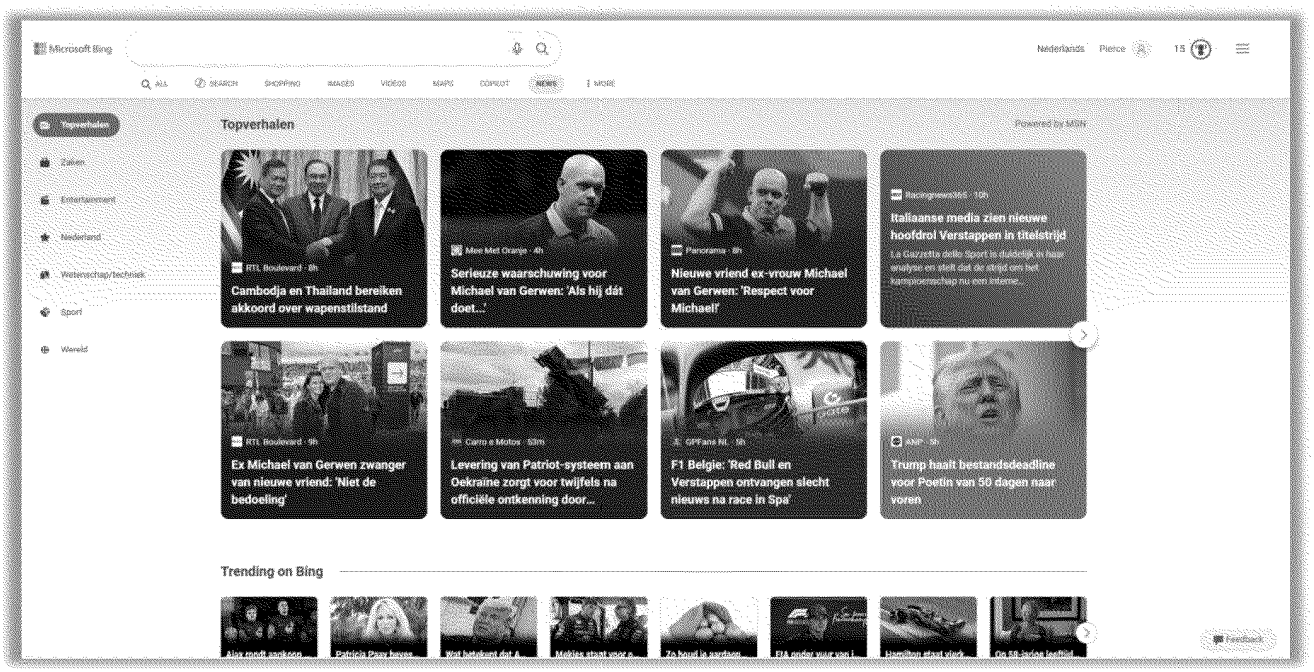
Figure 9: Bing Shopping landing page screenshot



News

Bing News is a dedicated news vertical which delivers high authority news content for Bing users, including news search results, news answers, and news carousels. While Bing News relies on the same main parameters for ranking results as Bing, content that appears in Bing news search results must meet the Bing News Publisher Guidelines, described in the Bing News Publication Hub. News sites interested in being included in the Bing News index may apply through the Publication Hub.

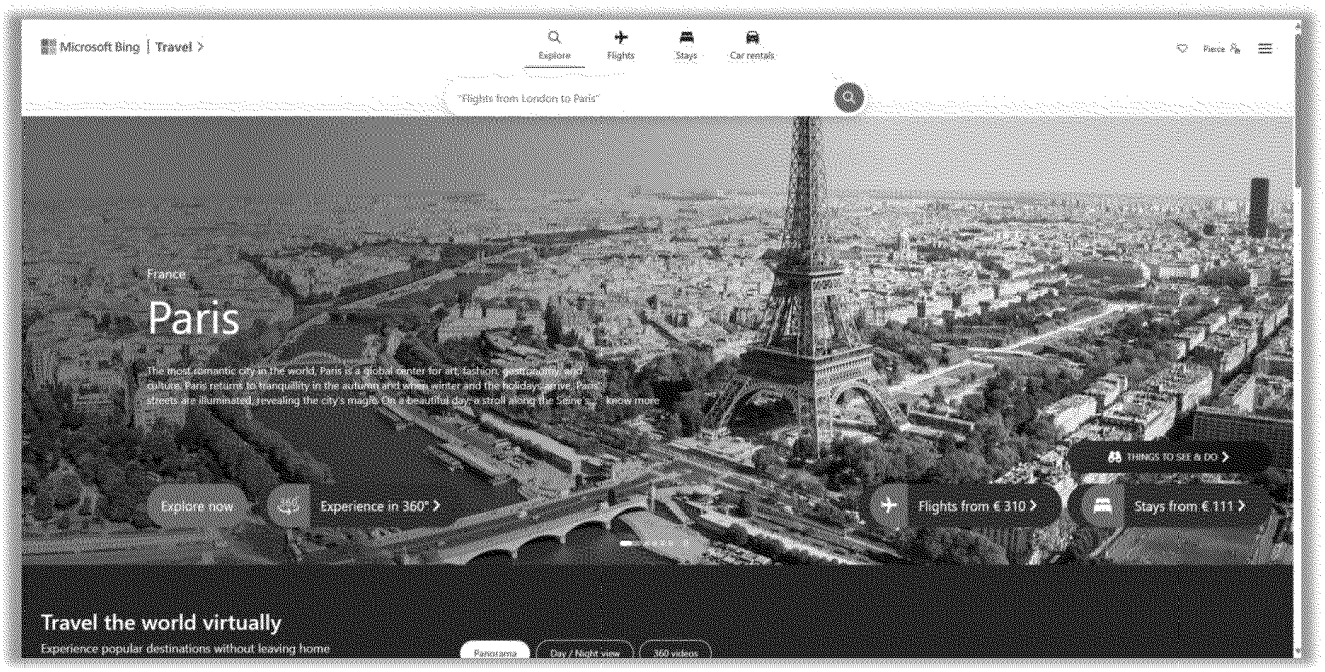
Figure 10: Bing News search vertical



Travel

Bing Travel is a specialized search and trip-planning tool that helps users find information about travel destinations, flights, accommodation, car rentals, and travel packages offered on third party websites. Users can search for flights, hotels, and other travel information. While Bing may display pricing and other key details, users complete all bookings through external booking websites either via links to third party travel websites or through embedded booking tools operated by third party travel partners. Bing does not manage the payment or booking transactions. Bing Travel may also suggest activities under "What to see & do" for the desired trip and promote other travel-related content. Bing Travel also includes "Travel Stories" for select locations, featuring third-party photo and video content curated for Bing by a third-party content partner.

Figure 11: Bing Travel home page screenshot



Advertisements Appearing on Bing

Bing, as a free service, is monetized almost entirely through advertising. When a user conducts a search on Bing, the user is shown both "organic" (i.e., unpaid) results and, when relevant, advertisements (i.e., paid advertising). Advertisements across Bing are clearly labeled as "Ads" and are displayed to users primarily based on their search queries. Advertisements that appear on Bing are provided by Microsoft Advertising, which is a business-to-business service separate from Bing that delivers advertising to a range of Microsoft services as well as third parties. As a separate service, advertisements on Bing are governed by the Microsoft Advertising Agreement, including the Microsoft Advertising Network Policies.

Ancillary Search Features Risk Profile

Some ways that systemic risks can manifest on ancillary search features absent sufficient mitigations include:

Maps

- Risk that Maps could provide incorrect information related to polling locations.

- Risk that Maps show content that perpetuate bias, reinforce stereotypes, or discriminate against products, vendors, or businesses.
- Risk that Maps recommends sites and destinations where illegal, dangerous, or regulated goods or services can be acquired.
- Risk that Maps' aerial and street level imaging of sensitive locations, including military bases, may expose national security assets.
- Risk that images of children which do not conceal their privacy are shown on the "Travel the World Virtually" or in the Maps features.

Shopping

- Risk that Shopping over moderates and removes available products that are related to certain political parties or figures in a manner that perpetuates political bias.
- Risk that users could input targeted query seeking products in Shopping with slogans or graphics that include discriminatory content and view such products in the first few pages of results.
- Risk that Shopping shows recommended products that include graphics or logos that are violent, gory, vulgar, or profane.
- Risk that Shopping results link to websites related to the sale of illegal, dangerous, infringing, or regulated goods.
- Risk that Shopping listings show products with logos or graphics that promote/include gender-based violence.
- Risk that activities on Shopping are used to collect funds used to finance terrorist organizations.

News

- Risk that News returns low authority or low-quality news sources with election misinformation.
- Risk that third party web content linked in News results includes hate speech or discriminatory content or content that perpetuates hateful attitudes or harmful stereotypes or otherwise facilitates discrimination towards individuals or groups.
- Risk that in order to comply with government orders, News may block access to third party web content which could result in unnecessarily censoring legitimate access to information.
- Risk that News results include content that is sexual, pornographic, or promote legal prostitution.
- Risk that News shows content that instructs on, promotes, normalizes, or praises illegal materials.
- Risk that News results show content that content that instructs on, promotes, normalizes, praises, or facilitates promotes suicide, self-harm, or causing harm to others.
- Risk that News shows content containing inaccurate or misleading information about public health issues, medical conditions, treatments, side effects, public health orders or decrees, or other information that could potentially undermine public health.
- Risk that News results include misinformation related to crisis events or issues of general public security.

Travel

- Risk that Travel provides insufficient information about travel product/services sellers, preventing informed consumer decisions.
- Risk that Travel shows locations, listings, or destinations that are used for consensual, legal prostitution.

- Risk that Travel ideas, itineraries, and/or search results suggest dangerous routes, destinations, or activities that place individuals in physical danger or otherwise undermine their physical well-being.
- Risk that private property is featured as a destination or things to do in Travel results.
- Risk that Travel offerings, such as coupons, expose users to phishing, hacking, or data breaches.
- Risk that Travel shows locations or rental listings that are targeted for terrorist training, funding, or support for terrorist acts.

Advertisements on Bing

- Risk that political advertisements target vulnerable Bing users with election-related misinformation or exhibit a lack of transparency in political advertising targeting practices that can mislead voters.
- Risk that Advertisements include content that misleads consumers about the legitimacy of businesses, products, or services.
- Risk that Advertisements on Bing are biased in their targeting, affecting protected groups' ability to access critical services.
- Risk that sexual or pornographic ads appear on Bing.
- Risk that Advertisements promote illegal content or activities.
- Risk that Advertisements appear on Bing that promote/show addictive or harmful products or illegal substances that can negatively impact mental or physical wellbeing.
- Risk that Advertisements on Bing are used to promote legal but harmful substances.
- Risk that Advertisements on Bing expose minors to harmful activities targeted for child sexual exploitation and abuse.

The key mitigation measures Bing has implemented to address these potential risk manifestations are described in Bing's Approach to User Safety and in Appendix II: Bing's Digital Safety Risk and Compliance Framework and Mitigations.

Changes to Bing Features

Bing endeavors to provide innovative experiences for its users and regularly evaluates the performance and usefulness of its features. As a result, features may be modified, tested, removed, or adjusted as Bing seeks to improve users' experiences. Throughout the year, Bing tracks updates to features, including rollbacks, changes, and launches through its pre-launch risk assessments, RAI Review, and Shiproom processes.

In addition to these core business processes, Bing Compliance Oversight and Defensive Intelligence Response and Transparency ("DIRT") teams have implemented an additional process to identify product launches that may be in scope for a Critical Impact Risk Assessment ("CIRA"), which evaluates potential impact of the launch of new feature or feature update on Bing's DSA systemic risk profile.

Feature updates since the 2024 Systemic Risk Assessment report are discussed below.

Microsoft Copilot Consumer Chat and Copilot Search

As noted above, Bing is continually innovating and evolving the experiences to delight its users. The evolution of Bing's Generative AI ("GenAI") offerings since the 2024 DSA Systemic Risk Assessment represent the dynamism of search and Generative AI. Changes occurred in October 2024, when the consumer chat experience, known as Bing Chat, was separated from the Bing service. Bing Chat, as well as broader conversational AI stack, became a standalone experience called Microsoft Copilot Consumer Chat,

available at copilot.microsoft.com. Bing also launched Bing Generative Search in October 2024, which was rebranded as Copilot Search in April 2025.

Copilot Search, Bing's generative AI search experience, gives users quick, summarized answers with cited sources and suggestions for further exploration, making it easier than ever to discover more. Copilot Search enhances users' search engine results page using GenAI, creating a curated experience tailored to users' queries. Copilot Search is grounded with real-time insights, and defaults to authoritative search results for topics that are not supported.

Copilot Search also incorporates enhanced safeguards, including policies limiting GenAI responses on sensitive topics, real-time classifiers that flag and filter problematic or sensitive information, and embedded RAI instructions via meta prompting that refuse to answer harmful or illegal queries.

Copilot Search includes mechanisms to monitor risk and assess the effectiveness of its safeguards. User feedback is collected in-product and regularly reviewed by a designated team to identify and address emerging issues. Bing also maintains social listening channels to identify quality or safety issues in public discourse.

This integration of GenAI within the Bing suite of services is aligned with Bing's *Trustworthy Search Principles* to ensure the careful balancing of users' fundamental right to access information with important societal values for privacy, security, and safety.

Bing Video Creator

Bing Video Creator, launched in late May 2025, is a similar product to Bing Image Creator in that it generates videos without audio based on the prompt the user inputs. This feature builds on the success of Bing Image Creator and reflects Bing's continued investment in multimodal GenAI capabilities. Bing Video Creator is designed to enhance user engagement and creative expression by offering a visual storytelling format that can be used across education, content creation, and communication contexts.

While the feature is still in early stages, it introduces new content generation dynamics that are monitored closely through existing safety frameworks. This includes alignment with Bing's RAI policies, prompt moderation, and integration with content filtering systems to mitigate the risks of misuse, such as the generation of misleading or sensitive visual material. By expanding into video content, Bing is broadening the scope of its generative service while maintaining a focus on safe and responsible deployment.

Real Estate

Bing officially deprecated its global Real Estate vertical at the beginning of 2025, as part of a broader strategic realignment aimed at consolidating efforts around core search functionalities and emerging GenAI capabilities.

Systemic Risk Assessment Methodology

Methodology Overview

Bing's Systemic Risk Assessment methodology was designed using a variety of available scientific and technical insights and included consultation with civil society organizations representing vulnerable users of Bing products and services.

As part of the Systemic Risk Assessment, the Bing Risk Assessment team considered the Probability and Severity of Inherent Risks due to the functioning, use, or misuse of Bing's products and services, identified implemented mitigations, and assessed whether those mitigations are effective, reasonable, and proportionate to the identified risks.

The Bing Risk Assessment team considered possible risk factors and influencers, such as the intentional manipulation of the service, the design of recommender systems and content moderation systems, and regional and linguistic aspects. The process that the Bing Risk Assessment team followed to conduct this period's Systemic Risk Assessment is described here along with a description of the enhancements made in the design of this period's assessment. Additional detail, including definitions and rating scales, is included in the Appendix I: Detailed Risk Assessment and Scoring Methodology.

Risk Assessment Process

Risk Definition and Discovery

The Bing Risk Assessment team identified twelve risk areas aligned with the risks identified in Article 34 of the DSA for consideration of the potential impact of the use or misuse of Bing's products and services.

1. Civic Discourse and Electoral Processes
2. Consumer Protection and Fraud
3. Discrimination and Hate
4. Freedom of Expression and Information
5. Human Dignity
6. Illegal Content and Activities
7. Mental and Physical Wellbeing
8. Private and Family Life
9. Protection of Personal Data
10. Public Health
11. Public Security
12. Rights and Protections of Minors

The Bing Risk Assessment team collected multi-disciplinary stakeholder input and information on updates to Bing features since July 2024, on the presentation and prevalence of risks across features, and on the implemented mitigations and ongoing monitoring activities across industry-standard best practices both in writing, through stakeholder engagements, and through a review of existing Microsoft documentation and data.

The Bing Risk Assessment team conducted workshops and interviews with key internal stakeholders across the organization to deep dive into the presentation of risks and unique mitigation measures across features. The Bing Risk Assessment team reviewed inputs from Microsoft's civil society engagements conducted

throughout the year and feedback from other external stakeholders to discuss areas of concern and best practices for mitigation.

The Bing Risk Assessment team gathered key internal metrics related to Bing policy enforcement; deconstructed transparency reporting metrics; and reviewed open-source information related to trends, patterns, and potential systemic risk related to use or misuse of Bing services, including Bing's social listening metrics and insights.

The Bing Risk Assessment team also examined authoritative sources and case studies, including Centre on Regulation in Europe ("CERRE") and European Commission Guidelines, on the systemic risk areas, collected relevant public policies and publications, and summarized relevant internal controls.

Inputs to the Systemic Risk Assessment:

1. Bing stakeholder validation of mitigations.
2. Guidance from authoritative sources, including the European Commission, on the severity of systemic risks.
3. Collection of Bing's published policies and official communications.
4. Bing internal stakeholder summarized policies and procedures.
5. Notes from internal consultations.
6. Notes from external consultations.
7. Bing internal metrics.
8. Transparency Reporting metrics.
9. Open-source data on public discourse related to Bing and systemic risk areas.
10. Bing products and features reviewed through the CIRA process
11. DSA risk and control matrix.

These inputs are described in more detail in Appendix I: Detailed Risk Assessment and Scoring Methodology.

Risk Analysis

The Bing Risk Assessment team assessed the Probability and Severity of each of the twelve systemic risks occurring on Bing through the potential use, misuse, or functioning of its products and services, as well as the maturity of risk mitigation measures implemented, to arrive at a prioritization of systemic risk areas with potentially higher levels of Residual Risk. The assessment incorporated the referenced inputs from the definition and discovery phase to inform ratings and implemented a traditional risk assessment equation to aid in prioritization of risk for additional enhancements to mitigation measures.

The Bing Risk Assessment team also considered for each applicable risk area and mitigation whether and how the following factors influence any of the systemic risks: the design of recommender systems and other relevant algorithmic systems; content moderation systems; applicable terms and conditions and their enforcement; systems for selecting and presenting advertisements; data-related practices; linguistic and cultural considerations; intentional manipulation including inauthentic use or automated exploitation; and amplification and potentially rapid and wide dissemination of illegal or violative content.

Probability

The Bing Risk Assessment team identified key risk scenarios associated with each risk area and considered the potential theoretical manifestation of each risk scenario across various features of the Bing service absent safeguards.

The Bing Risk Assessment team employed a data-driven probability assessment, combining insights from internal, external, and open-source data to assess the likelihood of specific risks stemming from use or misuse of Bing considering the vulnerability of the service absent mitigation measures and user demand or likely frequency of attempts to perpetrate each harm on the service. This methodology follows models used in climate and energy impact assessments.

Bing monitors a number of metrics in identifying and measuring risk across the service. Given that each metric has its own strengths and weaknesses, and thus cannot measure all aspects of a particular risk's potential negative impacts to Bing's services, the Bing Risk Assessment team incorporated several internal and external data points, including social listening data, survey data, DDR for the DSA systemic harm areas, and content moderation metrics, to inform the relative probability of the risk.

The Bing Risk Assessment team highlighted DDR for the DSA systemic harm areas in the individual risk analysis sections below, as Bing teams track and monitor these metrics to measure overall product health and evaluate the effectiveness of Bing's multi-layered safety system.

- DDR, including DDR for the DSA systemic harm areas, measures the likelihood that certain types of queries might result in harmful content showing in the top search results.
- Bing uses "adversarial sets," sampling queries that are focused on high-risk topics to help zoom in on the areas prone to showing harmful content and help better identify gaps and drive targeted improvements. Due to the nature of the "adversarial sets," DDR focuses on a very small portion of high-risk queries, over indexing on risky queries over non-risky queries in order to allow for more nuanced risk tracking.
- Bing uses a "diluted" DDR to represent the raw DDR of DSA systemic risks across overall search traffic, while also considering the trigger rate of the ***defensive search interventions***. The Defensive classifier trigger rate is <1%, meaning that across all user queries, less than 1% of queries might potentially lead to problematic content.

Severity

The Bing Risk Assessment team continues to focus on objectivity, using available data as part of a system-based assessment of severity. This methodology follows models used in environmental impact assessments and entails examining the complex systems impacted by the systemic risk area (geographic, political, security, environmental, societal, and wellbeing), while considering the scale (individual to global) and gravity (or ability to remediate) of impact to arrive at a severity score and rating.

Risks with irremediable impact on a greater number of systems on a broader scale will receive the highest severity ratings. For example, risks related to Consumer Protection and Fraud are rated as High (not Critical) severity because, while some impacts may be irremediable, the greatest impact is at the individual level and primarily on economic systems. Risks related to Public Security are rated as Critical as this risk area includes a number of risks with irremediable impacts (loss of life, for example) across multiple systems (security, political, and wellbeing) that reverberate across the local, country, and potentially regional levels.

Maturity of mitigations

The Bing Risk Assessment team organized existing mitigations and controls according to industry-standard best practices and evaluated their implementation according to the Digital Trust & Safety Partnership ("DTSP") Maturity Rating Scale. The DTSP Maturity Rating Scale is a best practice framework that uses a scale from "Ad Hoc" to "Optimized." The Bing Risk Assessment team used this scale to support the evaluation of the reasonableness, proportionality, and effectiveness of Bing's existing mitigations to the Probability and

Severity of the considered risks and risk manifestations and identified potential areas for enhancements to existing mitigations or new mitigations to further reduce the potential impact of the considered systemic risks.

Scoring

The Bing Risk Assessment team calculated the Inherent Risk for each risk area resulting from combining the ratings of Probability and Severity. The team then reduced the Inherent Risk rating by a percentage proportional to the assessed strength of the relevant mitigations. This methodology produces a view of Residual Risk (i.e., the remaining risk after accounting for applicable mitigations) across the risk areas to enable a more objective measurement of risk and to identify risks or mitigations that may warrant further investment, in alignment with Article 35 of the DSA.

For a more detailed description of the risk calculation, please refer to Appendix I: Detailed Risk Assessment and Scoring Methodology.

Assessment Tooling

Risk Assessment Workbook

The Bing Systemic Risk Assessment Workbook is the primary analytical tool supporting Bing's 2025 Systemic Risk Assessment Report. It is designed to provide a structured, transparent, and comprehensive view of systemic risks across the Bing product ecosystem. The workbook integrates quantitative scoring methodologies with qualitative risk narratives to assess how specific risk scenarios may manifest across products, the likelihood and severity of those risks, and the effectiveness of existing mitigations.

Organized into results, methodology, assessment, and scoring components, the workbook includes detailed registers of risks, products, and mitigations, as well as scoring worksheets that calculate risk ratings using standardized scales.

Risk Assessment Results

Overall Results

This section of the Report outlines the Residual Risk remaining within each Risk Area after the application of Bing's implemented mitigations to the identified Inherent Risks. For each Risk Area, the Bing Risk Assessment team has defined the risk areas and included theoretical ways they may manifest on the service, described the calculation of and rationale for the Inherent Probability and Severity ratings, and included the key mitigations relevant to that Risk Area. The rating scales and definitions are included in Appendix I: Detailed Risk Assessment and Scoring Methodology section of the Report.

The Risk Areas are presented in this section in the order of highest Residual Risk, followed by highest Inherent Probability, and then by lowest Maturity of Mitigations.

In the 2025 DSA Systemic Risk Assessment, the Residual Risk Ratings for Bing remain Low or Minimal across most assessed areas. However, the two risk areas, Illegal Content and Activities, and Human Dignity, are rated as Moderate. These areas continue to reflect an appropriate balance of fundamental rights and interests consistent with the unique risk profile of a search engine. While Bing assesses that current mitigations are effective, the service acknowledges the evolving nature of these risks and is committed to ongoing refinement of its mitigation strategies to maintain compliance with its systemic risk obligations under the DSA.

Figure 12: 2025 DSA Systemic Risk Assessment Ratings

Risk Area	Inherent Probability	Inherent Severity	Inherent Risk Rating	Mitigation Maturity	Residual Risk Rating
Freedom of Expression and Information	Remote	High	Low	Optimized	Minimal
Human Dignity	Likely	Critical	High	Managed	Moderate
Illegal Content and Activities	Likely	Critical	High	Managed	Moderate
Protection of Personal Data	Highly Likely	High	High	Optimized	Low
Civic Discourse and Electoral Processes	Likely	Critical	High	Optimized	Low
Consumer Protection and Fraud	Likely	High	High	Managed	Low
Discrimination and Hate	Not Likely	High	Moderate	Managed	Low
Mental and Physical Wellbeing	Not Likely	Critical	Moderate	Defined	Low
Private and Family Life	Likely	High	High	Defined	Moderate
Public Health	Not Likely	Critical	Moderate	Managed	Low
Public Security	Not Likely	Critical	Moderate	Managed	Low
Rights and Protections of Minors	Not Likely	Critical	Moderate	Defined	Low

Changes to Bing's Risk Profile

Over the past year, Bing has continued to operate steadily within a dynamic and evolving risk environment of technological, geopolitical, and social changes. Its risk profile remains stable even as new technologies emerge, and digital threats continue to evolve. As a VLOSE, its core function to enable users to access high quality information has not fundamentally changed, even with the integration of GenAI to make that more efficient.

This consistency within Bing's ecosystem, combined with the steady application of safety and relevance principles, contributes to a stable systemic risk profile. While not indicative of a change in the risk profile itself, Bing has expanded its risk capabilities by monitoring DDR aligned with DSA systemic risk areas. This enhancement enables more tailored analysis and reporting of harm-specific trends and mitigations over time.

One area of notable investment has been Bing's response to the growing concern around image-based abuse and the non-consensual sexual exploitation of individuals. Bing developed new NCII policies that apply across its ecosystem, as well as interventions to improve its abilities to detect and manage potential NCII, including classifiers to identify harmful prompts and potentially harmful content. Bing Image Creator downloads the StopNCII initiative's database of PhotoDNA hashes for reported and verified NCII content and continuously scans and removes NCII content from Bing's image index. Bing also updated the report a concern interface to include specific NCII reporting. Even with these investments, Bing recognizes the need and is continuing to explore ways to improve its ability to manage the complex risks associated with NCII.

Bing also implemented safeguards against prompts related to grooming and the generation of sexualized content involving children. These efforts are part of an evolving framework that incorporates human review, clear policy enforcement, and low-latency user appeals.

Bing has also made investments in the layered defenses for its Bing's GenAI capabilities, including bolstering upstream filters to detect jailbreak attempts, midstream blocking of known high-risk prompts, and downstream controls that restrict the generation of content related to public figures or harmful subjects. These targeted mitigations are complemented by continuous measurement pipelines. Bing uses both representative (realistic distribution of risky and non-risky queries) and adversarial (focused only on risky queries) datasets, as well as real-time social media monitoring, to detect emerging abuse patterns. These safeguards are regularly refined to ensure they remain effective without compromising user access to information.

As the service continues to evolve, Bing assesses how changes may affect systemic risks. The evolution of Copilot Search has been accompanied by careful evaluation of potential risk factors, with product design focused on respecting user rights to freedom of access to information as well as providing high authority content through grounding. Many core search functions remain unchanged, even with further enhancements to generative capabilities, and overall risk levels have not materially shifted. Bing continues to monitor external risk signals and internal safety metrics to ensure its mitigation strategy remains proportionate, resilient, and aligned with user rights and expectations.

Emerging Trends and Shifts in Global Risk Posture

Various events around the world can also impact Bing's risk profile, including geopolitical, health, and climate events as well as emerging technologies and trending social movements. As part of the 2025

Systemic Risk Assessment, the Bing Risk Assessment team considered a variety of relatively higher risk events to determine whether each warranted a shift in the Inherent Risk rating.

These relatively higher risk events include continued geopolitical unrest and active armed conflicts and the continued expansion of the use of GenAI capabilities for both malicious and beneficial purposes.

The environment for VLOSEs has also changed over the past year. Most notably, there has been a rise in NCII, a decrease in public concerns in the EU around health misinformation, particularly compared to the heightened concern during the COVID-19 pandemic, and a relative decline around elections-based risks following the 2024 "year of elections". Risks related to violent extremism have remained relatively steady, though the groups and focal points continue to evolve. Bing has observed and is aware of these shifts and continues to monitor how they may impact its ecosystem. That said, Bing has not identified any significant impact on its own risk profile as a result of these broader global changes.

Over the past year, there have been multiple high visibility instances of AI-generated deepfakes often involving lewd portrayals or portrayals related to politics and elections. Bing is aware of the risks related to both AI generated deepfakes surfacing in Bing search results, as well as Bing users using Bing Image and Video Creator to create deepfakes.

Bing has implemented safeguards, such as C2PA enabled signing and the Content Credentials tool to help review and verify the authenticity of digital content. This system is part of a broader initiative to combat inauthentic content and ensure transparency in content creation and editing. Bing also implements the system of mitigations for Bing Gen AI experiences as discussed above. With respect to EU specific risks, Bing has not observed risks disproportionality affecting vulnerable or EU-based users.

Due to the dynamic and ever-evolving nature of online safety, there have been some shifts in the global risk

Changes in Risk Rating

Figure 13: Changes from 2024 to 2025 Systemic Risk Assessment Ratings

Risk Area	Residual Risk Rating
Freedom of Expression and Information	**Minimal -**
Human Dignity	Moderate
Illegal Content and Activities	**Moderate +**
Private and Family Life	**Moderate +**
Protection of Personal Data	Low
Civic Discourse and Electoral Processes	Low
Consumer Protection and Fraud	Low
Discrimination and Hate	Low
Mental and Physical Wellbeing	**Low -**
Public Health	Low
Public Security	**Low -**
Rights and Protection of Minors	**Low -**

Key: + indicates an increase of a rating from last year's assessment by one rating level, - indicates a decrease of a rating by one rating level

For Residual Risk Rating, there was a reduction from four Moderates in 2024 to two Moderates in 2025. From the 2024 to 2025 SRA, there are three reasons for changes in the risk ratings, (1) changes that are clerical in nature, (2) changes in the information environment, and (3) changes resulting from enhancements to Bing's mitigation measures.

With respect to clerical changes, the Illegal Content and Activities Residual Risk Rating increased from Low to Moderate over the past year primarily due to the internal reassignment of risks related to CSAM from the Rights and Protection of Minors risk area to the Illegal Content and Activities risk area. The Residual Risk Rating assigned to Rights and Protection of Minors also consequently decreased from Moderate to Low. This change better aligns the risk with legal obligations and enforcement priorities. Also, the Residual Risk Rating for Mental and Physical Wellbeing has decreased from Moderate to Low. This shift is based on internal and external indicators showing a lower likelihood of systemic harm. This is partly due to the passage of time since the launch of Gen AI outputs, which, while initially generating heightened scrutiny, have since become more normalized, and thus less likely. Bing recognizes that concerns remain and continues to invest in mitigation measures to safeguard and support user wellbeing considering an ever-changing risk area and product offerings.

Due to changes in the information environment in the EU and the world, the Residual Risk Rating for Public Security decreased from Moderate to Low based on a lower probability rating for Public Security in 2025. While still present within the EU information environment, public discourse related to the Israel-Gaza War has decreased in frequency since the 2024 assessment period, which impacted the Public Security probability rating. While risks related to terrorism typically require a longer time horizon than one year to materially decrease, they may swiftly and unexpectedly increase. Therefore, Bing is not changing its approach toward mitigating risks related to terrorism and continues to make investments in relevant mitigations.

Bing adjusted the Freedom of Expression and Information Residual Risk rating from Low to Minimal for this assessment. A significant factor in this shift is the removal of Chat from Bing and the creation of a separate Microsoft Copilot Consumer Chat product, as well as the Generative Search experience, recently rebranded as Copilot Search, that falls back to core search for complex or potentially harmful queries.

The rating changed because Copilot Search prioritizes providing users access to high quality information and, for queries that potentially present a higher risk of unexpected exposure to harmful content, defaults to providing authoritative search results rather than providing generative responses. This is in contrast with Microsoft Copilot Consumer Chat which, part of which was formerly Bing Chat, which blocks the generative responses for queries that present a higher risk of unexpected exposure to harmful content. Bing also continues to invest in monitoring overblocking for Bing Image and Video Creator results, to continually strive for more consistent and fair experiences.

The remaining risk ratings are consistent with the prior year assessment, which in some cases also reflects enhancement of mitigation measures to address risks that increased in probability during this assessment period. Bing invested in mitigation measures related to NCII across products, but industry developments and public concern related to NCII increased in the past year, therefore Bing assessed Private and Family Life to be a Moderate Residual Risk.

Results per Risk Area

Freedom of Expression and Information

One of the Trustworthy Search principles that guides Bing is to promote free and open access to information within the bounds of the law, with respect for local law and other fundamental rights. This is reflected in its careful efforts to not unduly restrict important interests such as freedom of expression, open access to information, and media pluralism. Given that search engines are the primary way users find information online and conduct research relevant to their daily and professional lives, over-moderation of content in search could have a significant negative impact on the right to access information and freedom of expression. Bing must carefully balance these competing fundamental rights and interests as it works to ensure that its algorithms return provide high quality, relevant content in response to the user's queries without unduly limiting their ability to access information or express themselves.

Bing endeavors to provide open access to as much of the web as possible, but in limited cases may undertake certain interventions (such as removal of a website or downranking) such as where the content violates local law, or Microsoft's policies. When limiting access to content, Bing strives to ensure its actions are narrowly tailored so fundamental rights are not unduly restricted. In some cases, different features may require different interventions based on functionality and user expectations. In most cases, rather than removing or blocking content, Bing informs users of certain risks through public service announcements or warnings or provide users with options for tailoring their content.

Following Bing's principles, Bing's safety systems are balanced with the rights to Freedom of Expression and Information. Bing provides transparency around Bing's safety practices, content moderation processes and content removal data. For GenAI features, although Microsoft may take more pre-emptive action to limit generation of certain types of harmful content, it carefully measures and monitors overblocking rates (i.e. rates at which responses are restricted) to drive improvements.

As previously mentioned within the *Changes in Risk Rating* section, considering the Inherent Probability, Inherent Severity, and the Maturity of Mitigations relevant to Freedom of Expression and Information, the Bing Risk Assessment team has assessed the Residual Risk related to Freedom of Expression and Information on the service to be "Minimal"(See *Changes in Risk Rating* section for further details).

Risk Area	Inherent Probability	Inherent Severity	Inherent Risk Rating	Mitigation Maturity	Residual Risk Rating
Freedom of Expression and Information	Remote	High	Low	Optimized	Minimal

Risk Definition

Risk Area	Freedom of Expression and Information
Risk Definition	Risk that content or activities with actual or foreseeable negative effects on the fundamental right to Freedom of Expression and Information, including the freedom and pluralism of the media occur on the service.
Risk Scenarios	• Over-moderation or overblocking of user-generated content online, over-removal, unjustified removal, or restriction of expressive content • Overly restricting access to information online, over-removal or restriction of access to lawful information

	• Recommendations that suppress pluralism • Inadequate internal safeguards against undue government or law enforcement influence over online content or safety interventions • Content or activities that chill public discourse, participation in online spaces, or open expression online • Content or activities that otherwise negatively impact freedom of expression, information, or pluralism
Theoretical Risk Manifestations absent Mitigations	• Risk that search over moderates the search results, removing or downranking URLs limiting access to information. • Risk that search's ranking processes disproportionately favor or demote certain types of content or viewpoints. • Risk that Bing's internal safety interventions could prevent individuals from accessing or generating information important to engaging in civic life, pursuing interests, or exercising other rights, or such safety interventions may be applied in biased ways. • Risk that Bing may unnecessarily reduce or block access to content based on government or law enforcement requests. • Risk that Shopping over moderates' available products (e.g., with slogans or graphics that may seem offensive). • Risk that News results might lead users to third party web content that is biased. • Risk that in order to comply with government orders, News may block access to third party web content which could result in unnecessarily censoring legitimate access to information. • Risk that Bing Image or Video Creator disproportionately block certain prompts or generated responses • Risk that the mitigations in Bing Image or Video Creator (such as blocklists, filtering, classifiers, metaprompts and other safeguards) disproportionately restrict certain types of content users seek • Risk that Bing Image or Video Creator disproportionately generates images that align with particular cultural standards or viewpoints, resulting in less diverse representations of the population • Risk that Bing Image or Video Creator limits artistic and journalistic expression by favoring certain prompts over others • Risk that Bing Image or Video Creator over moderation practices could lead to the prevention of image generation, even with lawful and fair prompts, thus restricting visual expression • Risk that GenAI Protection Systems flag and block the generation of legitimate content

Inherent Risk

The **Inherent Probability** of systemic risks related to Freedom of Expression and Information stemming from the use, misuse, or functioning of Bing's products and services absent sufficient mitigations is assessed as "Remote."

Bing considered internal metrics and social listening data as factors in assessing a "Remote" level of probability relative to other potential systemic risks on Bing.

Absent mitigations to ensure that Bing, in the application of its controls and safeguards, does not overly restrict freedom of expression or information, the likelihood of this occurrence is Remote as Bing generally does not remove third-party websites from the search index, absent legal or policy reasons.

The likelihood of Bing overly restricting freedom of expression or information on GenAI features is greater, but this is also balanced by the fact that users can continue to search for information using Bing web search (or

other search mediums), even where responses or outputs are restricted in GenAI features and that Copilot Search will return answers from core search if it cannot generate a GenAI response to the query.

The **Inherent Severity** of content and activities that negatively impact Freedom of Expression and Information on Bing is rated as "High" primarily due to the gravity of risks within this risk area, considering the potential for significant harm to individuals' fundamental rights as well as broader societal, media, and political systems that built upon the rights to free expression, access to information, and media plurality. Consequences related to the negative impact on Freedom of Expression and Information can be irremediable at scale. But the impact of many risks within this risk area is not considered significant.

With the rating for Inherent Probability as "Remote" and Inherent Severity as "High," the **Inherent Risk** associated with Freedom of Expression and Information on the Bing service is "Low."

Maturity of Mitigations

Bing has implemented an expansive set of mitigations to address risks related to Freedom of Expression and Information, including those described in Bing's Approach to User Safety, in Appendix II: Bing's Digital Safety Risk and Compliance Framework and Mitigations, and those risk-specific mitigations summarized below.

The mitigations Bing has implemented follow best practices that promote trust and safety in every aspect. As such, the **Maturity of Mitigation** efforts Bing has applied to the Freedom of Expression and Information risk area is assessed as "Optimized," which brings the **Residual Risk** Rating down to "Minimal."

Nevertheless, the rights to free expression and access to information remain fundamental to Bing's approach to online safety. Bing continues to invest in innovating and enhancing strategies to further mitigate and manage risks related to Freedom of Expression and Information, particularly with respect to GenAI features that offer new ways for users to approach research, learning, and search. The highlights of the key mitigations currently implemented are described below.

Risk Specific Mitigations

The following section highlights both how broader service-wide mitigations described in Bing's Approach to User Safety and inventoried in Appendix II, apply for this risk area, and include mitigations that are uniquely tailored to address the identified risks in this area.

The mitigations described here include both unique interventions specific to Freedom of Expression and Information and expansions on previously described cross-platform mitigations with specific applications to Freedom of Expression and Information. Where applicable, Bing's broader infrastructure has been adapted to address the particular risks associated freedom of access to information. This layered approach ensures both platform-wide consistency and harm-specific responsiveness. For ***prioritizing high authority content in search,*** central to Bing's strategy is a set of search principles that respects the Freedom of Expression and Information while balancing it with other key interests and the risks of harm occurring on its service. Bing continuously engages in internal and external discussions to evaluate and refine the principles.

As a search engine, the ***fundamental rights to access and seek information*** and freedom of expression are core to Bing's Trustworthy Search principles. Bing limits removal of content to narrow scenarios to avoid unduly impacting access to information. Bing provides training and oversight to its content review teams to ensure consistent application of policies, including reviews of decisions, and provides escalation paths to local legal experts as needed. For government demands, Bing employs additional safeguards to ensure any actions taken are narrow, specific, submitted in writing, and based on valid legal orders. Bing has automated safeguards in place to trigger additional reviews as needed. Where content is removed from search, Bing is

transparent by notifying users through a notice on the search engine results page and publishing regular transparency reports (available on Reports Hub | Microsoft CSR).

Bing's ***Transparency Reports*** as applied to Freedom of Expression and information, further provide transparency on content moderation.

As an additional quantitative signal of safety performance, Bing ***monitors overblocking metrics*** (identifying where Bing Image or Video Creator has unnecessarily blocked a creative response) in order to continually to assess the effectiveness and appropriateness of its GenAI safety measures to appropriately prioritize users' fundamental rights to information.

Bing engages in regular self-directed tests and audits of its system to ensure its ability to respond and Microsoft, including Bing, also responds to periodic evaluations by the third-party independent organization ***Ranking Digital Rights*** which publishes annual ratings on Bing's practices, governance, and leadership on the protection of freedom of expression and privacy.

Microsoft also is a member of the ***Global Network Initiative ("GNI")***, a multistakeholder collaboration to protect freedom of expression and privacy in tech and submits to a regular audit by GNI to ensure Bing has sufficient protections in place to uphold free expression. Bing is biannually audited for its commitments to human rights, including free expression, as part of its membership in the GNI.

Human Dignity

Risks related to Human Dignity include exposure to content pertaining to sexual exploitation, human trafficking, vulgarity, and gore.

Bing's ranking algorithms are designed to protect users from unexpected exposure to content that could negatively affect users' fundamental rights to Human Dignity. Bing provides users with a SafeSearch option to control their search experience to filter out adult content and/or content including explicit language. GenAI features deploy additional mitigations to prevent harmful AI-generated outputs such as pornographic content or graphic violence. Considering the Inherent Probability, Inherent Severity, and the Maturity of Mitigations relevant to Human Dignity, the Bing Risk Assessment team has assessed the Residual Risk related to Human Dignity on the service to be "Moderate," consistent with the rating assigned in 2024.

Risk Area	Inherent Probability	Inherent Severity	Inherent Risk Rating	Mitigation Maturity	Residual Risk Rating
Human Dignity	Likely	Critical	High	Managed	Moderate

Risk Definition

Risk Area	Human Dignity
Risk Definition	Risk that content or activities with actual or foreseeable negative effects on the fundamental rights to Human Dignity occur on the service.
Theoretical Risk Scenarios covered by this Risk Area	• Human trafficking (other than sex trafficking) or coercive conduct depriving others of their human rights for personal gain • Sexual exploitation, non-consensual prostitution or pornography, or sex trafficking, and extreme pornography • Graphic violence, human gore, vulgarity, profanity, or other content that demeans, shocks, or offends • Sexual or pornographic content, and consensual, legal prostitution • Content or activities that otherwise negatively impact Human Dignity

Theoretical Risk Manifestations absent Mitigations	• Risk that search engines return content that facilitates or depicts human trafficking in the top search results. • Risk that search engines return content that instructs on, promotes, normalizes, or praises human trafficking other than sex trafficking in the top search results. • Risk that search engines return content that can be used to access the services of individuals being trafficked in the top search results. • Risk that search results or autosuggest features lead users to web results that aid human traffickers by promoting false job postings. • Risk that search engines return content that is graphic, violent, gory, vulgar, profane or that otherwise demeans, shocks, or offends in the top search results. • Risk that search engines return content that is sexual or pornographic or link to information on locating consensual and legal prostitution in the top search results • Risk that Maps or Travel show locations, listings or destinations that are used by malicious actors for human trafficking • Risk that Maps or Travel show locations, listings or destinations that are used by malicious actors for sexual exploitation or non-consensual prostitution • Risk that News results include content that is sexually exploitative, graphic, pornographic, violent, vulgar, or profane. • Risk that Shopping shows recommends products that include graphics or logos that are violent, gory, vulgar, or profane. • Risk that News results include content that is sexual, pornographic or promote legal prostitution. • Risk that sexual or pornographic ads appear on Bing. • Risk that Maps and Travel show locations, listings or destinations that are used for consensual, legal prostitution. • Risk that Bing Image or Video Creator is used to create materials that depict sexual exploitation, pornography, or extreme pornography • Risk that Bing Image or Video Creator is used to create violent, gory, or profane content • Risk that Bing Image or Video Creator is used to generate images that are used to recruit or exploit potential trafficking victims • Risk that Bing Image or Video Creator is used to generate or manipulate images that are sexually exploitative, contain pornographic content, or promote prostitution • Risk that Bing Image or Video Creator facilitates the generation of explicit adult content involving real or synthetic individuals without consent • Risk that Bing Image or Video Creator is used to generate manipulate images that are violent, gory, vulgar, or include profanity • Risk that GenAI summaries are based on low authoritative sources and produce content that may be harmful or misleading. • Risk that adversarial user queries or grounding data may bypass the safety features for Copilot Search and produce harmful or misleading content

Inherent Risk

The **Inherent Probability** of systemic risks related to Human Dignity stemming from the use, misuse, or functioning of Bing's products and services absent sufficient mitigations is assessed as "Likely."

Bing considered internal metrics and social listening data as key factors in assessing a “Likely” level of probability relative to other potential systemic risks on Bing.

The **Inherent Severity** of content and activities that negatively impact Human Dignity is rated as “Critical” primarily due to the gravity of risks within this risk area, considering the potential for significant harm to wellbeing, societal, economic, and security systems at the individual and potentially broader level. Some inherent risks within this risk area, such as human trafficking and prostitution, have potentially irremediable

impacts. Other risks, such as gore, vulgarity, profanity, or pornographic content, may be remediable but can be significant.

With the rating for Inherent Probability as "Likely" and Inherent Severity as "Critical," the **Inherent Risk** associated with negative impacts to Human Dignity on the Bing service is "High."

Maturity of Mitigations

Bing has implemented a robust set of mitigations to address risks related to Human Dignity, including the mitigations that apply across harms as described in Bing's Approach to User Safety, and inventoried in Appendix II: Bing's Digital Safety Risk and Compliance Framework and Mitigations, and well as risk-specific mitigations summarized below.

The mitigations Bing has implemented relative to Human Dignity follow best practices with defined, documented, and managed processes. As such, the **Maturity of Mitigation** efforts Bing has applied to the Human Dignity risk area is assessed as "Managed," which brings the **Residual Risk** Rating down to "Moderate," as detailed above.

While not a proxy for Residual Risk, Bing has measured the DDR for Human Dignity on Bing Search as an average of 1.28% from September 2024 to April 2025. This means that Bing estimates 1.28% of search traffic may include content that may have negative effects related to Human Dignity.

Bing nevertheless, continues to invest in developing and enhancing strategies to further mitigate and manage risks related to Human Dignity. The highlights of the key mitigations currently implemented are described below.

Risk Specific Mitigations

The following section highlights both how broader service-wide mitigations described in Bing's Approach to User Safety and inventoried in Appendix II, apply for this risk area, and include mitigations that are uniquely tailored to address the identified risks in this area.

The mitigations described here include both unique interventions specific to Human Dignity and expansions on previously described cross-platform mitigations with specific applications to Human Dignity. Where applicable, Bing's broader infrastructure has been adapted to address the particular risks associated with degrading, explicit, or exploitative content. This layered approach ensures both platform-wide consistency and harm-specific responsiveness.

To uphold the principle of Human Dignity, Bing has implemented ***safety measures for search suggestions*** aimed at curbing the generation of suggestions that might tarnish this fundamental right. This also includes creating avenues for users to effortlessly report suggestions they deem inappropriate, ensuring such concerns are promptly addressed.

As part of the ***Safety Measures – Defensive Search Interventions*** mitigation, Bing has a dedicated team accountable for implementing algorithmic defensive search interventions (as well as metrics monitoring and remediation) to address high impact issues in search results, such as information manipulation, hate speech, and other problematic content that could negatively impact user safety.

Bing uses query classifiers to identify queries that may be associated with elevated risk, such as those that could surface degrading, harmful, or exploitative content, and those trigger a set of defensive search interventions. These algorithmic measures are integrated rather than standalone and work together to uphold Bing's relevance, quality, and credibility standards—particularly in cases where harmful content may exploit data voids or manipulate user intent.

When a query is flagged as potentially risky, Bing may respond with measures such as boosting high-authority sources or applying algorithmic dampening to deprioritize low-quality or harmful content. These interventions are applied in accordance with Bing's core search principles and are actively monitored by the dedicated team that manages implementation and assesses impact across high-risk areas. This combined approach allows Bing to detect potentially harmful search patterns early and respond with targeted mitigations that preserve both user safety and access to high-quality information.

For ***safety features on visual search***, Bing also uses the hash-matching technologies PhotoDNA and MD5 to detect matches of previously identified CSEAI in these images provided by users to prevent upload. In the context of the search, the use of these technologies furthers Bing's goal to avoid surfacing potentially harmful web content to users. More broadly, images uploaded to Bing Visual Search typically contribute to training Bing's image matching algorithms. By scanning image upload, Bing helps to ensure that CSEAI is not included in training data.

Microsoft regularly engages in ***social listening for GenAI features*** to review the prevalence of issues within Bing Image or Video Creator to evaluate the landscape of potential harms to Human Dignity and adjust the safeguards accordingly. Through this process, Microsoft enhances the mitigations to reduce the occurrence of harmful outputs potentially impacting Human Dignity.

Users can use "***report a concern***" (See *Content Moderation* section for further details) to report unexpected offensive or harmful material (e.g., reporting unexpected adult, violent, or gore in results, non-consensual intimate imagery, or Harm - CSEAI).

The ***Terms and Conditions for Users***, found within the Microsoft Service Agreement (MSA), as applied to Human Dignity prohibit using Microsoft services to publicly display, generate, or share inappropriate content or material (involving, for example, nudity, bestiality, pornography, offensive language, or graphic violence). The Bing Image Creator Terms of Use further prohibit use of Bing's GenAI features to create or share adult content, violence or gore, or content that is otherwise disturbing or offensive.to create or share adult content, violence or gore, or content that is otherwise disturbing or offensive.

Illegal Content and Activities

Risks related to Illegal Content and Activities include CSEAI, potential IP infringement, defamation, and the promotion or sale of illegal, dangerous, or regulated goods or services.

Bing takes proactive steps to prevent the dissemination of CSEAI using PhotoDNA scanning. Bing's content moderation systems are designed to process requests regarding illegal content, with reviews conducted by the appropriate teams. While Bing employs automated content detection (ACD) to identify and exclude from the index certain high-risk content, most notably child sexual exploitation and abuse imagery (CSEAI) and, more recently, NCII, the majority of content removal decisions rely on a report-and-react model.

For most types of harmful or policy-violating content, automated detection is not feasible due to the high degree of contextual nuance involved. Instead, Bing responds to reports through established content moderation systems and service-level agreements (SLAs), in accordance with its notice-and-takedown framework. Bing includes information on illegal content removal in transparency reports.

Considering the Inherent Probability, Inherent Severity, and the Maturity of Mitigations relevant to Illegal Content and Activities, the Bing Risk Assessment team has assessed the Residual Risk related to Illegal Content an Activities on the service to be "Moderate" (Refer to *Changes in Risk Rating* above for further details).

Risk Area	Inherent Probability	Inherent Severity	Inherent Risk Rating	Mitigation Maturity	Residual Risk Rating
Illegal Content and Activities	Likely	Critical	High	Managed	Moderate

Risk Definition

Risk Area	Illegal Content and Activities
Risk Definition	Risk related to the conduct of illegal activity or dissemination of illegal content through the service.
Risk Scenarios	• IP infringement, defamation • Promotion, concealment, acquisition, or sale of illegal, dangerous, or regulated goods (including counterfeit goods, animals, drugs, psychoactive substances, criminal property, firearms, and other weapons) and services relevant to the same • Facilitation of unlawful immigration • Causing suffering of animals • Creation, dissemination, or promotion of CSEAI • Content or activities that otherwise disseminating illegal content or enabling illegal activity through the service
Theoretical Risk Manifestations absent Mitigations	• Risk that search engines return content that infringes on IP rights or disseminates defamatory statements in the top search results. • Risk that search engines return content that contains illegal content sources, such as third-party websites offering illegal content streaming or music downloads or sites providing counterfeit goods or other deceptive commercial items for sale in the top search results. • Risk that search engines return otherwise illegal content in the top search results. • Risk that search engines return content that includes CSAM or CSEAI or applications, tools, or services for creating synthetic CSEAI. • Risk that Shopping results show counterfeit or knock-off products that infringe on the IP of brands. • Risk that Shopping results link to websites related to the sale of illegal, dangerous, infringing, or regulated goods (including illegal drugs, weapons, psychoactive substances, pharmaceuticals, etc.) • Risk that Advertisements promote illegal content or activities. • Risk that Maps recommends sites and destinations where illegal, dangerous, or regulated goods or services can be acquired. • Risk that News shows content that instructs on, promotes, normalizes, or praises illegal materials. • Risk that Copilot Search or Bing Image or Video Creator are used to generate third party IP without authorization in generated responses • Risk that Bing Image or Video Creator are used to generate third party IP without authorization in generated responses • Risk that Bing Image or Video Creator is used to generate information or images that encourage animal cruelty • Risk that Bing Image or Video Creator is used to create images that promote or glorify illegal activities, products, or substances

Inherent Risk

The **Inherent Probability** of systemic risks related to Illegal Content and Activities appearing on deriving from use of Bing's products and services absent sufficient mitigations is assessed as "Likely".

Bing considered internal metrics and social listening data as key factors in assessing a "Likely" level of probability relative to other potential systemic risks on Bing.

The **Inherent Severity of** content and activities that negatively impact Illegal Content and Activities on Bing is rated as "Critical" primarily due to the gravity of risks within this risk area considering the potential for harm to security, environment, and wellbeing systems up to the regional and even the global level for environment systems. Some risks within this risk area have potentially irremediable consequences.

With the rating for Inherent Probability as "Likely" and Inherent Severity as "Critical," the **Inherent Risk** associated with Illegal Content and Activities on the Bing service is "High."

Maturity of Mitigations

Bing has implemented a robust set of mitigations to address risks related to Illegal Content and Activities, including those described in Bing's Approach to User Safety, in Appendix II: Bing's Digital Safety Risk and Compliance Framework and Mitigations, and those risk-specific mitigations summarized below.

The mitigations Bing has implemented relative to Illegal Content and Activities follow best practices with defined, documented, and managed processes. As such, the **Maturity of Mitigation** efforts Bing has applied to the Illegal Content and Activities risk area is assessed as "Managed," which brings the **Residual Risk** rating down to "Moderate."

While not a proxy for Residual Risk, Bing has measured the DDR for Illegal Content and Activities on Bing Search as an average of .14% from September 2024 to April 2025. This means that Bing estimates .14% of search traffic may include content that may have negative effects related to Illegal Content and Activities.

Nevertheless, Bing continues to invest in developing and enhancing strategies to further mitigate and manage risks related to Illegal Content and Activities. The highlights of the key mitigations currently implemented are described below.

Risk Specific Mitigations

The following section highlights both how broader service-wide mitigations described in Bing's Approach to User Safety and inventoried in Appendix II, apply for this risk area, and include mitigations that are uniquely tailored to address the identified risks in this area.

The mitigations described here include both unique interventions specific to Illegal Content and Activities and expansions on previously described cross-platform mitigations with specific applications to Illegal Content and Activities. Where applicable, Bing's broader infrastructure has been adapted to address the particular risks associated with illegal content. This layered approach ensures both platform-wide consistency and harm-specific responsiveness.

In response to user search queries for CSAM, Bing displays ***in-product indicators*** and PSAs stating that CSAM is illegal; Bing also includes links to report these illegal materials or receive anonymous support, aimed at those who are victims of CSAM. For image and video queries relating to CSAM, no results are shown and instead Search displays unrelated images from popular web content. For Bing Image and Video Creator, the response to CSAM related queries is disengagement stating that the request is inappropriate or harmful.

Similarly, Bing also states in its policies that advertisers who are misusing Microsoft services in a way that could cause harm to users or the service, such as advertising in fraudulent, harmful, or illegal manner are ***subject to immediate account suspension***.

For ***incident response and SLAs***, Bing has a robust reporting and reactive response infrastructure that allows it to quickly action notices of illegal materials and other policy violating content (including exposed

personal information, CSEAI, etc.) from governments, users, or other stakeholders. Support teams are trained in applicable requirements and policies and provide escalation paths where needed for complex issues, including to local legal experts.

Users are encouraged to ***report advertisements*** that promote illegal content or violate Microsoft Advertising terms through a structured feedback form available on Bing's pages. Specific forms for reporting low quality ads or IP concerns are also available.

The ***Report a Concern*** forms include functionality to report unlawful content.

To further bolster the fight against illegal content, Bing leverages both internal and ***authoritative external sources***, such as the IWF and the International Social Service ("ISS"), to enhance its ranking and relevance capabilities. This approach is particularly crucial in identifying and demoting potentially illegal materials, where, for instance, a high volume of valid copyright infringement notices can signal a site's low-quality, influencing its ranking negatively.

The ***Terms and Conditions for Users***, found within the Microsoft Service Agreement (MSA), as applied to Illegal Content and Activities prohibit using Microsoft services to do anything illegal, or try to generate or share content that is illegal; to publicly display, generate, or share inappropriate content or material (involving, for example, criminal activity); and to violate or infringe upon the rights of others (e.g., unauthorized sharing of copyrighted music or other copyrighted material). The Bing Image Creator Terms of Use further prohibit use of Bing's GenAI features to infringe on others' legal rights, including intellectual property rights; to create or share child sexual exploitation or abuse material; or to do anything illegal.

Protection of Personal Data

Risks related to the Protection of Personal Data include exposure to content pertaining to collection, processing, release of data, targeted ads, hacking, malware, phishing, data breaches, insufficient protection of data, and unauthorized disclosure or exposure of personal data.

Bing adheres to Microsoft privacy standards and relevant laws. Bing completes Data Protection Impact Assessments ("DPIA") for any new products or features that involve materially different personal data processing. Bing also adheres to Microsoft-standard security practices that protect stored data from inappropriate access, via both technical and organizational means. Bing limits the retention of data to that necessary to provide the service.

Microsoft's longstanding belief that privacy is a fundamental human right has also informed every stage of Bing's development and deployment of the Copilot Search experience. Microsoft's commitment to protecting the privacy of users, including by providing individuals with transparency and control over their data and integrating privacy by design through data minimization and purpose limitation, are foundational to Copilot Search. As Bing evolves its approach to providing the Copilot Search GenAI experiences, it will continually explore how better to protect privacy. More information about how Microsoft protects its users' privacy is available in the Microsoft Privacy Statement. Considering the Inherent Probability, Inherent Severity, and the Maturity of Mitigations relevant to the Protection of Personal Data, the Bing Risk Assessment team has assessed the Residual Risk related to the Protection of Personal Data on the service to be "Low," and furthermore, did not see a change in the scoring for Protection of Personal Data for since 2024.

Risk Area	Inherent Probability	Inherent Severity	Inherent Risk Rating	Mitigation Maturity	Residual Risk Rating

Protection of Personal Data	Highly Likely	High	High	Optimized	Low

Risk Definition

Risk Area	Protection of Personal Data
Risk Definition	Risk that content or activities with actual or foreseeable negative effects on the Protection of Personal Data occur on the service.
Risk Scenarios	• Collection, processing, access to, or release of information without consent or other misuse of data, including data mining or targeted advertisements without consent • Phishing, hacking, degradation or other abuse of service or data breaches or other unauthorized access, alteration, destruction, or disclosure of data or otherwise insufficient protection of data • Disclosure or exposure of personal information or data online • Other content or activities negatively impacting the Protection of Personal Data
Theoretical Risk Manifestations absent Mitigations	• Risk that users are harmed by inappropriate processing of personal data, such as lack of transparency, control, access to remedies, or disclosures. • Risk that information requested to be removed under a "Right to be Forgotten" request is incorrectly processed or technical issues impact removal of content subject to a valid removal request. • Risk that search history data, if inadvertently disclosed, reveals an individual's beliefs, relationships, sexuality, medical issues, or other private information. • Risk that search engines return content that enables easy retrieval of information about an individual that falls within their private sphere and that the individual did not intend or consent to make available publicly, such as extremely sensitive personal information that could create risks of identity thefts, including credit card numbers, medical records, etc. in the top search results. • Risk that users of Ancillary Features may be harmed by inappropriate processing of personal data they've provided to the service. • Risk that Travel offerings, such as coupons, expose users to phishing, hacking, or data breaches. • Risk that personal information shared with Ancillary Features could be inadvertently disclosed or exposed online. • Risk that Bing Image or Video Creator is used to generate images that are used for phishing campaigns

Inherent Risk

The **Inherent Probability** of systemic risks related to the Protection of Personal Data stemming from the use, misuse, or functioning of Bing's products and services absent sufficient mitigations is assessed as "Highly Likely."

Bing considered internal metrics and social listening data as key factors in assessing a "Highly Likely" level of probability relative to other potential systemic risks on Bing. Public discourse reflects a higher level of concern related to the security of data provided to GenAI tools, which leads to the assessment of inherent probability as "Highly Likely." The **Inherent Severity** of content and activities that negatively impact the Protection of Personal Data is rated as "High" primarily due to the gravity of risks within this risk area, considering the potential for significant harm to economic, security, and societal systems at the individual and local level. Some risks within this risk area – particularly related to phishing, hacking, malware, or data breaches – may reach a broader scale of impact while others, such as collection of data without consent, may only impact the individual. For the most part, these risks are remediable but may include some irremediable damage.

With the rating for Inherent Probability as "Highly Likely" and Inherent Severity as "High," the **Inherent Risk** associated with the Protection of Personal Data on the Bing service is "High."

Maturity of Mitigations

Bing has implemented an expansive set of mitigations to address risks related to the Protection of Personal Data, including those described in Bing's Approach to User Safety, in Appendix II: Bing's Digital Safety Risk and Compliance Framework and Mitigations, and those risk-specific mitigations summarized below.

The mitigations Bing has implemented relative to the Protection of Personal Data follow best practices promoting Trust and Safety in every aspect. As such, the **Maturity of Mitigation** efforts Bing has applied to the Protection of Personal Data risk area is assessed as "Optimized," which brings the **Residual Risk** rating down to "Low."

While not a proxy for Residual Risk, Bing has measured the DDR for Protection of Personal Data on Bing Search as an average of .02% from September 2024 to April 2025. This means that Bing estimates .02% of search traffic may include content that may have negative effects related to Protection of Personal Data.

Nevertheless, Bing continues to invest in innovating and enhancing strategies to further mitigate and manage risks related to the Protection of Personal Data. The highlights of the key mitigations currently implemented are described below.

Risk Specific Mitigations

The following section highlights both how broader service-wide mitigations described in Bing's Approach to User Safety and inventoried in Appendix II, apply for this risk area, and include mitigations that are uniquely tailored to address the identified risks in this area.

The mitigations described here include both unique interventions specific to the Protection of Personal Data and expansions on previously described cross-platform mitigations with specific applications to Protection of Personal Data. Where applicable, Bing's broader infrastructure has been adapted to address the particular risks associated with unauthorized disclosure, phishing, or hacking. This layered approach ensures both platform-wide consistency and harm-specific responsiveness.

Bing Data retention is managed within a set schedule. Bing limits the retention of data to that necessary to provide the service and ensure the safety of the service. Unauthenticated users can control their search history through Bing's in-product controls using Bing's settings page.

Copilot Search has data retention and deletion policies to help ensure that personal data collected through this feature is kept as long as needed.

Bing was built with privacy in mind, so that personal data is collected and used as needed and is retained no longer than is necessary. Microsoft applies industry-leading data storage policies and practices to ensure the ***safety and security of user data*** that is collected and stored. Some of these practices include, but are not limited to, storing data in Microsoft-owned and operated data centers with standard physical security access and the data is encrypted from data subject to data center in transit.

The Visual Search feature deploys a ***blurring mechanism*** that blurs faces in the images at the time of the upload by users, so that facial images are not further processed or stored. Access to most Bing Search history data is limited to Bing employees. More information about the personal data that Bing collects, how it is used, and how it is stored and deleted is available in the Microsoft Privacy Statement.

For any new products or features that pose materially different personal data processing challenges, Bing completes or updates ***DPIAs*** as part of its rigorous privacy protocol. These assessments are reviewed by

Microsoft's Data Protection Officer and are subject to reevaluation annually or sooner, should there be a significant change to processing risks. This process ensures that Bing's products and services comprehensively identify and address privacy risks at each level of development and deployment.

Bing's ***Report a Concern*** tool includes options for users to report exposed personal or private information.

Bing's established "***Right to Be Forgotten***" policies and procedures help reduce risks of negative effects to Protection of Personal Data by providing reporting tools for users to submit requests that certain online content associated with them be removed from Bing.

Civic Discourse and Electoral Processes

Risks related to Civic Discourse and Electoral Processes include exposure to content containing election misinformation, risks third party content indexed in search or generated in AI features could lead to political polarization, and targeted efforts by bad actors to leverage technology to undermine civic institutions and democratic elections.

Search engines like Bing play a key role in ensuring users have access to information related to Civic Discourse and Electoral Processes. Balancing user safety with freedom of information and expression and avoiding the introduction of bias or influence is critical to this risk area. Bing takes this responsibility seriously and has implemented robust measures to promote election integrity and reduce the risk of systemic harm related to Civic Discourse and Electoral Processes on Bing.

Global elections are highly complex, as each country has different election structures, voting processes, timing, political systems, and party structures. During the Assessment Period, there were elections in some European Union member states. Mitigating risks related to Civic Discourse and Electoral Processes on the Bing service presents challenges in the scale and speed of electoral process, developments surrounding elections, parties, and candidates, and threats posed by malicious actors.

This complexity is furthered by the continued rise in GenAI accessible online. In recognition of these novel risks and harms, in November 2023 Microsoft announced a set of Election Protection Commitments grounded in four principles to help safeguard voters, candidates and campaigns, and election authorities worldwide. These principles, which help inform Bing's response and safety program, are:

- Voters have a right to transparent and authoritative information regarding elections.
- Candidates should be able to assert when content originates from their campaign and have recourse when their likeness or content is distorted by AI for the purpose of deceiving the public during the course of an election.
- Political campaigns should protect themselves from cyber threats and be able to navigate AI with access to affordable and easily deployed tools, trainings, and support.
- Election authorities should be able to ensure a secure and resilient election process and have access to tools and services that enable this process.

Bing takes a multifaceted approach to protecting Civic Discourse and Electoral Processes and regularly updates its policies, and practices to adapt to evolving risks, trends, and technological innovations, as well as regulatory expectations. Bing partners closely with dedicated Microsoft internal teams focused on elections, including Democracy Forward and specialized cross-functional teams dedicated to elections and specific risks arising from AI & Elections.

In addition to Bing's core ranking algorithms and systems designed to promote high authority content, Bing deploys numerous product and process improvements consistent with DSA Election Guidelines. Bing's core

ranking algorithms and systems include ingestion of election and misinformation-related threat intelligence narratives and localized election data to inform product interventions, direction of users to official sources and high authority content, partnerships with internal and external experts to collect insights on emerging threats and trends to inform safety efforts, display of special "How to Vote" answers pointing to authoritative local sources ahead of specific elections, rapid incident response processes, coordination (through Microsoft) with election authorities, and investment in GenAI mitigations and improvements intended to address risks of hallucination or inaccurate information related to elections and political deepfakes.

Considering the Inherent Probability, Inherent Severity, and the Maturity of Mitigations relevant to Civic Discourse and Electoral Processes, the Bing Risk Assessment team has assessed the Residual Risk related to Civic Discourse and Electoral Processes on the service to be "Low," and furthermore did not see a change in the scoring for Civic Discourse and Electoral Processes for this year.

Risk Area	Inherent Probability	Inherent Severity	Inherent Risk Rating	Mitigation Maturity	Residual Risk Rating
Civic Discourse and Electoral Processes	Likely	Critical	High	Optimized	Low

Risk Definition

Risk Area	Civic Discourse and Electoral Processes
Risk Definition	Risk that content or activities with actual or foreseeable negative effects on Civic Discourse and Electoral Processes occur on the service.
Risk Scenarios	• Coordinated manipulation, manipulation, disinformation, inauthentic activity, misleading communications, digital threats, and foreign interference, including deceptive GenAI election content • Election misinformation • Political polarization, political bias, and the creation of echo chambers • Intimidation, harassment, coordinated targeting, or coercion of political or civic actors • Content or activities otherwise negatively impacting Civic Discourse and/or Electoral Processes
Theoretical Risk Manifestations absent Mitigations	• Risk that bad actors manipulate search results or exploit data voids to surface low authority, low quality content with an intent to deceive or manipulate audiences. • Risk that search engines return content that is low authority, low quality, outdated, or other content that includes inaccurate, misleading information about electoral processes, voting procedures, or election outcomes in EU member states and around the world in the top search results. • Risk that lack of authoritative information or data voids in less trafficked languages or geographies leads to lower quality search results in certain markets • Risk that Bing's safety systems introduce bias or limit access to important election information if applied inconsistently. • Risk that search results or autosuggest features might lead users to web results that favor certain political views, influencing perceptions and polarizing suggestions related to elections. • Risk that Bing's partnerships with external experts or fact checkers introduce bias into the Bing service. • Risk that News results include disinformation, manipulated content, inauthentic content, content produced by malign foreign actors related to elections, electoral processes, voting procedures, election outcomes in EU member states and around the world; or content providing efforts to undermine election integrity. • Risk that News returns low authority or low-quality news sources with election misinformation.

	• Risk that political advertisements target vulnerable Bing users with election-related misinformation or exhibit a lack of transparency in political advertising targeting practices that can mislead voters. • Risk that Maps could provide incorrect information related to polling locations. • Risk that News returns news sources that increase political polarization and filter bubbles. • Risk that Shopping over moderates and removes available products that are related to certain political parties or figures in a manner that perpetuates political bias. • Risk that News inadvertently returns news articles that are targeted to damage a political actor's reputation or credibility. • Risk that bad actors misuse Bing Image and Video Creator to generate deceptive election content, such as deceptive images of political candidates, elections, and voting processes • Risk Bing Image and Video Creator exacerbates societal divisions through the creation of images that might exaggerate political stances, deepening echo chambers • Risk that bad actors misuse Bing Image and Video Creator is to generate images that are used to intimidate, harass, or coerce political or civic actors into complying with their demands • Risk that GenAI summaries are based on low authoritative sources and produce content that may be harmful or misleading • Risk that adversarial user queries or grounding data may bypass the safety features for Copilot Search and produce harmful or misleading content

Inherent Risk

The **Inherent Probability** of systemic risks related to Civic Discourse and Electoral Processes stemming from the use, misuse, or functioning of the Bing service absent sufficient mitigations is assessed "Likely."

Bing considered internal metrics and social listening data as key factors in assessing a “Likely” level of probability relative to other potential systemic risks on Bing.

The Bing Risk Assessment team also considered the number of elections in Europe over the assessment period as a potential impact to the probability level. The Inherent Probability rating is lower this year due to the lower number of elections in Europe compared to last year.

The **Inherent Severity** for content and activities that negatively impact Civic Discourse and Electoral Processes on Bing is rated as “Critical” primarily due to the gravity of risks within this risk area, considering the potential for significant harm to political, societal, economic, and security systems at the local, country, and even regional levels. Some risks within this risk area have potentially irremediable consequences on political or security systems. The impact of other risks may be considered remediable but still has the potential to cause significant damage.

With the rating for Inherent Probability as “Likely” and Inherent Severity as “Critical,” the **Inherent Risk** associated with Civic Discourse and Electoral Processes on the Bing service is “High.”

Maturity of Mitigations

Bing has implemented an expansive set of mitigations to address risks related to Civic Discourse and Electoral Processes, including those described in Bing’s Approach to User Safety, in Appendix II: Bing’s Digital Safety Risk and Compliance Framework and Mitigations, and those risk-specific mitigations summarized below.

The mitigations Bing has implemented relative to Civic Discourse and Electoral Processes follow best practices that promote trust and safety in every aspect. As such, the **Maturity of Mitigation** efforts Bing has

applied to the Civic Discourse and Electoral Processes risk area is assessed as "Optimized," which brings the **Residual Risk** Rating down to "Low."

As part of preparing for elections in the EU, Bing reviewed and implemented the guidance provided by the European Commission (Commission Guidelines for providers of Very Large Online Platforms and Very Large Online Search Engines on the mitigation of systemic risks for electoral processes pursuant to Article 35(3) of Regulation (EU) 2022/2065). Bing's existing practices reflected the guidelines, building on the additional improvements made in 2024, the "year of elections" in the EU. Microsoft and Bing have also undertaken multiple engagements with the Commission, press agencies, and Digital Service Coordinators.

While not a proxy for Residual Risk, Bing has measured the DDR for Civic Discourse and Electoral Processes on Bing Search as an average of .03% from September 2024 to April 2025. This means that Bing estimates .03% of search traffic may include content that may have negative effects related to Civic Discourse and Electoral Processes.

Nevertheless, Bing continues to invest in innovating and enhancing strategies to further mitigate and manage risks related to Civic Discourse and Electoral Processes. The highlights of the key mitigations currently implemented are described below.

Risk Specific Mitigations

The following section highlights both how broader service-wide mitigations described in Bing's Approach to User Safety and inventoried in Appendix II, apply for this risk area, and include mitigations that are uniquely tailored to address the identified risks in this area.

Where applicable, Bing's broader infrastructure has been adapted to address the particular risks associated with elections, content integrity, and information manipulation. This layered approach ensures both platform-wide consistency and harm-specific responsiveness.

Bing takes a multi-layered approach to protecting election integrity, with mitigations across the various layers of the Bing Safety Funnel. Bing also regularly updates its strategies and practices to adapt to evolving risks, trends, and technological innovations, as well as new election announcements (many of which occur on short notice) and regulatory guidance.

Product Quality

The ***Terms and Conditions for Users***, found within the Bing Image Creator Terms, as applied to Civic Discourse and Electoral Processes, prohibit use of the service to engage in activity that is false or misleading; to attempt to create or share content that could mislead or deceive others, including for example creation of disinformation, or deceptive impersonation; to remove or alter content credentials or other provenance methods, marks, or signals that indicate that the Creations were generated by the service; or to otherwise attempt to mislead others about whether the Creations were generated by the service.

Furthermore, Microsoft ***does not permit political advertising*** within the Microsoft Advertising ecosystem, which includes Bing. Advertising policies prohibit ads for political candidates, parties, ballot measures, fundraising, or other election-related content globally. Ads that were issue-based and could be perceived as politically sensitive are also disallowed. Microsoft Advertising is a COPD signatory, and its specific commitments are outlined in Microsoft's most recent COPD report.

Microsoft's Content Integrity Tools, originally designed as a pilot program for the 2024 election cycle and to gain feedback about Content Credentials-enabled tools, allow users to add content credentials to their own authentic content, in effect ***Content Credentials as a Service***. The Tools were available to political

campaigns in the EU, as well as to elections authorities and select news media organizations in the EU and globally.

These tools included a partnership and collaboration with fellow Tech Accord signatory, TruePic. Announced in April 2024, this collaboration leveraged TruePic's mobile camera SDK enabling campaign, election, and media participants to capture authentic images, videos and audio directly from a vetted and secure device. Called the "Content Integrity Capture App" (an app that makes it easy to directly capture images with C2PA enabled signing) launched for both Android and Apple and can be used by participants in the Content Integrity Tools pilot program.

Product Safety

Bing has ***special informational and answer segments for*** **elections** including 'How to Vote' answers in defined markets to provide users with high-authority information related to the election, localized by language and member state. Bing Search directs users to official election sources, such as https://elections.europa.eu/, and other high-authority sites locally.

Beyond standard measurement to evaluate the quality of search results against our policies, and ahead of specific elections, Bing ***monitors elections-related DDR metrics*** through election-related query sets to measure the presence of content that goes against its search policies across Bing Search features including Web, Image, and Video. Bing may invest additional resources into evaluating the quality of search results through sample-based metric evaluations and 'red team' testing on Core and Copilot Search and/or creative content generation features.

Bing undertakes post-election reviews, as appropriate, to evaluate product and mitigation performance, reflect on challenges and learnings, and identify potential areas for improvement. These reviews took place both in product review settings and in broader cross-functional teams dedicated to elections at Microsoft.

Targeted Interventions

Microsoft identifies, tracks, and reports potential information operations and foreign malign influence campaigns, working with independent third-party partners and through the ***threat intelligence team focused on nation state actors and information operations***. Public reporting from that team on cyber influence operations can be found on their website at Microsoft's Threat Intelligence Team (MTAC).

Copilot Search and Bing Image and Video Creator utilize defensive interventions as well as local election mitigations, such as blocklists, to restrict generation of images or certain types of content concerning politicians in upcoming elections and to prevent responses to certain election-related queries out of an abundance of caution. Through Microsoft's dedicated cross-functional team focused on democracy and elections, these defensive interventions and election-related mitigations include information on political parties, candidates, and elections from local election authorities (including in the EU) and other high-authority third-party sources. Bing continues to invest in these mitigations, including through focused testing in key EU languages used by Bing users in the EU.

Bing uses ***social listening*** to evaluate online conversations around potential RAI risks related to election and political topics, enabling mitigation actions as necessary.

Content Moderation

Microsoft maintains an "***Election Communications Hub***" to support democratic governments and political parties worldwide as they build secure and resilient election processes. This hub gives election authorities the ability to rapidly report any issues or concerns to Microsoft security and support teams in the days and weeks leading up to their elections, allowing swift assistance with major security challenges.

Microsoft maintains a web page for ***deceptive AI election reporting***, where political candidates can report concerns about deepfakes on Microsoft services. This tool empowers political candidates globally to aid in detecting harmful deepfakes. The webpage is available and localized throughout the EU.

Research and Partnership

Microsoft undertakes significant efforts to support election integrity through a whole-of-company approach. Several cross-functional teams, including dedicated internal product reviewers, global and regional election working groups, and specialists in elections and democracy focus on election-related risks, policies, and AI-related issues. These teams regularly review product features and inform decisions across Bing Search, GenAI features, and other relevant services.

Microsoft has a global election working group that convenes stakeholders from Bing, local support teams, and broader Microsoft policy functions to share insights, identify risks, and coordinate mitigations related to civic integrity worldwide. These cross-functional ***teams at Microsoft dedicated expressly to addressing election-related concerns*** ensure Microsoft undertakes efforts to improve its processes to support election integrity using a whole-of-company approach.

A key component of the dedicated cross-functional team's strategy is establishing lines of communication between election authorities in EU member states and ***identifying risks to electoral processes,*** including potential foreign information operations targeting elections.

Microsoft maintains a "***Campaign Success Team***" to help political campaigns navigate cybersecurity challenges and the evolving landscape of AI. This team advises and supports campaigns in combating cyber-influence operations and protecting the authenticity of their content and images.

Microsoft implements ***political group awareness campaigns and trainings***, such as Deceptive AI and Elections Trainings for political parties and campaigns across the EU. These trainings help participants learn to identify deceptive uses of AI and misinformation in elections and report them using Microsoft's Deceptive AI Reporting Tool (discussed below).

Despite evolving threats to democracy, Microsoft continues to protect open and secure democratic processes by providing services and technologies to secure critical institutions, ***safeguard electoral processes from cyberattacks***, and build public trust in voting procedures.

Microsoft works closely with election-related customers to provide ***security guidance*** and tools to improve cyber-resilience and protect electoral integrity.

Microsoft supports the International Foundation for Electoral Systems ("IFES") in strengthening the cybersecurity practices of investigative journalists reporting on abuses of state resources in elections. Microsoft also partnered with the National Democratic Institute ("NDI") to strengthen cybersecurity infrastructure for political parties and campaigns internationally. Microsoft engages consistently with ***civil society and industry partners*** through a dedicated democracy and elections team. These partnerships included collaboration with International IDEA and the NATO Hybrid Centre of Excellence to identify and promote best practices for supporting election integrity.

For ***external engagements with governments***, Bing and Microsoft representatives presented information on mitigations related to elections, both before and after EU elections, to the European Commission, digital service coordinators, and other member state authorities and has participated in EU elections tabletop workshops organized by the Commission dedicated to assessing risk scenarios concerning elections in the EU.

Additionally, Bing collaborates with Microsoft Research and external research organizations to support studies on safe product design, responsible AI, and information manipulation. In preparation for global elections, Microsoft Research conducts internal research on information integrity and elections in the era of GenAI.

Transparency and Reporting

Bing is a signatory to the ***EU COPD*** and as a signatory to the EU COPD, and as an active participant in the related working groups, Bing fulfills its commitments to reduce disinformation risks. This includes participation in the Elections Working Group's rapid response program, delivering transparency reporting, offering the Content Credential tool to help users assess the trustworthiness of websites or domains, and empowering them to make informed decisions about online content. Bing also provides enhanced reporting on elections designated as a "crisis" under the COPD framework.

As part of its commitments under the EU Code of Practice on Disinformation (COPD), Bing provides ***transparency reporting*** via biannual reporting under the Code on the systems and policies in place to prevent the spread of misinformation, including related to elections, across the EU. The biannual COPD reports include comprehensive information on measures related to declared crises and agreed-upon EU elections. Bing also serves as a joint coordinator in the Generative AI (GenAI) working group under the COPD.

Consumer Protection and Fraud

Risks related to Consumer Protection and Fraud include exposure to content pertaining to scams, spam, false representation and information, fraudulent businesses, bots, and deceptive commercial practices.

Bing's ranking algorithms are designed to prevent users from being exposed to low quality content, including websites engaging in scams or fraudulent activities. Bing's spam policies and detection protect users from web spam that manipulates search algorithms and does not add content value.

Bing recognizes that risks associated with fraud in advertising have been a major focus for regulators and platform providers like Bing in the last year, with concerns about how GenAI exacerbates these risks top of mind.

Considering the Inherent Probability, Inherent Severity, and the Maturity of Mitigations relevant to Consumer Protection and Fraud, the Bing Risk Assessment team has assessed the Residual Risk related to Consumer Protection and Fraud on the service to be "Low" (See below for further details).

Risk Area	Inherent Probability	Inherent Severity	Inherent Risk Rating	Mitigation Maturity	Residual Risk Rating
Consumer Protection and Fraud	Likely	High	High	Managed	Low

Risk Definition

Risk Area	Consumer Protection and Fraud
Risk Definition	Risk that content or activities with actual or foreseeable negative impact to a high-level of consumer protection occur on the service.
Risk Scenarios	• Scamming, defrauding, spamming, use of malware, virus, or spyware, fraudulent business or advertising, or identity theft • False representation, abuse of position, falsifying information, including the creation of fake accounts, use of bots, or other deceptive commercial practices, including false claims and misleading advertisements

	• Insufficient information about sellers or products, preventing informed consumer decisions • Content or activities that otherwise negatively impact consumer protection
Theoretical Risk Manifestations absent Mitigations	• Risk that search results or autosuggest features lead users to web results that inadvertently prioritize scam businesses or fraudulent websites. • Risk that search engines return content that is misleading to consumers about the legitimacy of businesses, products, or services in the top search results. • Risk that users encounter third-party content including instructions on how to defraud users in the top search results. • Risk that Advertisements link to malware. • Risk that Advertisement include content that misleads consumers about the legitimacy of businesses, products, or services. • Risk that Shopping and Travel listings include fraudulent, deceptive, or unsafe products. • Risk that Maps or News results include content that misleads consumers about the legitimacy of businesses, products, or services. • Risk that Travel provides insufficient information about travel product/services sellers, preventing informed consumer decisions. • Risk that Bing Image and Video Creator is misused to generate highly realistic images of non-existent individuals or impersonate companies for use in phishing schemes, fake endorsements, or efforts to create synthetic identities that could be used to defraud businesses and consumers • Risk that Bing Image and Video Creator is misused to insert virus or malicious code into generated images to create malware that can be used for malicious activities • Risk that GenAI summaries are based on low authoritative sources and produce content that may be harmful or misleading. • Risk that adversarial user queries or grounding data may bypass the safety features for Copilot Search and produce harmful or misleading content

Inherent Risk

The **Inherent Probability** of systemic risks related to Consumer Protection and Fraud stemming from the use, misuse, or functioning of Bing's products and services absent sufficient mitigations is assessed as "Likely."

Bing considered internal metrics and social listening data as key factors in assessing a “Likely” level of probability relative to other potential systemic risks on Bing.

The **Inherent Severity** for fraudulent content or activities on the Bing service is rated as "High" due to the gravity of risks within this risk area and the potential for significant economic impact at the individual-level. While economic, and even security, societal, and wellbeing, impact at the individual-level can be severe, the risk is generally limited to the individual-level and does not necessarily scale to the country, regional, or global impact scale. Not all risks within this risk area would be considered "High," as the impact of spam and deceptive advertisements are less likely to cause significant harm, compared to malware links or deceptive job postings.

With the rating for Inherent Probability as “Likely” and Inherent Severity as “High,” the **Inherent Risk** associated with Consumer Protection and Fraud on the Bing service is “High.”

Maturity of Mitigations

Bing has implemented a robust set of mitigations to address risks related to Consumer Protection and Fraud, including those described in Bing’s Approach to User Safety, in Appendix II: Bing’s Digital Safety Risk and Compliance Framework and Mitigations, and those risk-specific mitigations summarized below.

The mitigations Bing has implemented relative to Consumer Protection and Fraud follow best practices with defined, documented, and managed processes. As such, the **Maturity of Mitigation** efforts Bing has applied to the Consumer Protection and Fraud risk area is assessed as "Managed," which brings the **Residual Risk** rating down to "Low."

While not a proxy for Residual Risk, Bing has measured the DDR for Consumer Protection and Fraud on Bing Search as an average of .02% from September 2024 to April 2025. This means that Bing estimates .02% of search traffic may include content that may have negative effects related to Consumer Protection and Fraud.

Nevertheless, Bing continues to invest in developing and enhancing strategies to further mitigate and manage risks related to Consumer Protection and Fraud. The highlights of the key mitigations currently implemented are described below.

Risk Specific Mitigations

The following section highlights both how broader service-wide mitigations described in Bing's Approach to User Safety and inventoried in Appendix II, apply for this risk area, and include mitigations that are uniquely tailored to address the identified risks in this area.

The mitigations described here include both unique interventions specific to Consumer Protection and Fraud and expansions on previously described cross-platform mitigations with specific applications to Consumer Protection and Fraud. Where applicable, Bing's broader infrastructure has been adapted to address the particular risks associated with scams, fraud, and false representation. This layered approach ensures both platform-wide consistency and harm-specific responsiveness.

Spam tactics are often used by bad actors to manipulate the ranking and show up in search results. Bing invests significant time and resources to ensure that users are provided with high quality content to avoid the possibly of its page crawlers and algorithms inadvertently returning results that could negatively impact consumers, such as spam or other fraudulent websites. Bing continuously monitors manipulation trends in high-risk areas and deploys mitigations to ensure that quality results are returned to users.

Bing's Webmaster Guidelines include details on prohibited practices, specific to spam and fraud, intended to manipulate or deceive the Bing Search algorithms.

Similarly, Bing also states in its policies that advertisers who are misusing Microsoft services in a way that could cause harm to users or the service, such as advertising in fraudulent, harmful, or illegal manner are subject to immediate ***account suspension***.

Products in Shopping, primarily leverage product advertisements from Microsoft Advertising. For ***shopping content moderation***, Microsoft reviews each listing added to the Microsoft Merchant Center product feed. Product listings that do not comply with Microsoft Advertising Policies are disapproved. Any merchant that violates Microsoft Advertising Policies will also be disapproved, in this case no products from the merchant will be visible. Bing's Shopping team conducts another layer of review on top of Microsoft Advertising when including products within Shopping search results.

For consumer protection with ***Microsoft Advertising Policies***, Bing's Webmaster Guidelines, anti-spam/manipulation policies, and sensitive information policies are in place to address the risks of materials that might impact consumer data protection and could be used for scams or fraud, such as pages containing account information or credit card numbers.

The ***Microsoft Advertising fraud detection system*** is multi-tiered, with checks based on main line ("ML") models, heuristic sweeps/rules, and manual reviews to prevent bad actors from coming onto the system.

Overall, Bing has models and rules that apply at the advertiser-level as well as those that apply to a particular ad.

- Some of the models that apply at the advertiser-level include:
    - Payment fraud detection: This model triggers when the advertiser adds a payment instrument. This model is owned by the commerce risk team and Bing consume its decision.
    - Purchase risk model: This model triggers upon billing the advertiser to determine whether to try to make the charge or block the advertiser. This helps to bring up-to-date information to determine payment risk.
    - Behavioral risk model: This model looks at various features of an advertiser's account or campaign(s) such as the ad text, keywords, budgets, account age, account structure, attributes of the payment instrument, individuals associated with the account, devices from which the advertiser is accessing the system etc. to determine the risk of an advertiser being fraudulent. This model runs when the advertiser makes any changes to their ad campaigns (ads/keywords/budgets, etc.).
    - LLM-based spoofing detection: One of the large categories of fraud is where advertisers pretend to be a legitimate site (e.g., trying to deceive users into believing that an ad leads to the official website of a bank). Bing has an LLM-based approach looking at advertiser demand to try and predict if advertiser is engaging in this behavior.

The ***Terms and Conditions for Users***, found within the Microsoft Service Agreement (MSA), as applied to Consumer Protection and Fraud prohibit using Microsoft services to send spam or engage in phishing, or try to generate or distribute malware or engage in activity that is fraudulent, false or misleading. The Bing Image Creator Terms of Use further prohibit use of Bing's GenAI features to engage in activity that is fraudulent, false, or misleading; to attempt to create or share content that could mislead or deceive others, including for example creation of content enabling fraud; or to remove or alter content credentials or other provenance methods, marks, or signals that indicate that the Creations were generated by the service.

Discrimination and Hate

Risks related to Discrimination and Hate include exposure to content containing bias, discriminatory content, and hate speech. Bing applies its comprehensive safety mitigations to the wide range of web content that poses risks to Discrimination and Hate.

By prioritizing high authority content, Bing's ranking algorithms are designed to protect users from being unexpectedly harmed by biases, hate speech, and other discriminatory content. GenAI features deploy additional mitigations to prevent harmful AI-generated outputs. Additionally, GenAI features undergo review prior to launch to identify the potential for harm, including harm related to Discrimination and Hate, under Microsoft's RAI program. Considering the Inherent Probability, Inherent Severity, and the Maturity of Mitigations relevant to Discrimination and Hate, the Bing Risk Assessment team has assessed the Residual Risk related to Discrimination and Hate on the service to be "Low."

Risk Area	Inherent Probability	Inherent Severity	Inherent Risk Rating	Mitigation Maturity	Residual Risk Rating
Discrimination and Hate	Not Likely	High	Moderate	Managed	Low

Risk Definition

Risk Area	Discrimination and Hate

Risk Definition	Risk that content or activities with actual or foreseeable negative effect on the fundamental right to non-discrimination occur on the service, impacting individuals and groups based on race, ethnicity, religion, gender, sexual orientation, or other protected traits, or reflecting, reinforcing, or perpetuating harmful stereotypes, biases, or inequalities.
Risk Scenarios	• Bias and discriminatory practices, features, safety interventions, or policies unfairly disadvantaging protected individuals or groups • Bias and discriminatory practices, user actions, activities, or behaviors negatively impacting protected individuals or groups • Hate speech and discriminatory content, content that insults, degrades, or dehumanizes protected individuals or groups, including hateful imagery • Content or activities that otherwise negatively impact the fundamental right to non-discrimination
Theoretical Risk Manifestations absent Mitigations	• Risk that Bing algorithmic systems contain bias that negatively impacts protected user groups on Bing. • Risk that safety features and interventions are designed or implemented without adequate linguistic or cultural understanding, resulting in impacts to some users because of protected traits. • Risk that Bing's content moderation practices are implemented unfairly resulting in biased or discriminatory search results. • Risk that search results autosuggest features lead users to web results that could expose users, including minors, to hateful ideologies, beliefs, or symbols. • Risk that search engines return content that contain hate speech or discriminatory content in the top search results. • Risk that search engines return content that includes recorded livestreams from hate-fueled violent attacks in the top search results. • Risk that search engines return content that provides hateful rhetoric or advocates the use of violence against people because of their race, or other protected traits in the top search results. • Risk that advertisements on Bing are biased in their targeting, affecting protected groups' ability to access critical services. • Risk that Shopping, Maps, and Travel show content that perpetuate bias, reinforce stereotypes, or discriminate against products, vendors, or businesses. • Risk that Shopping shows products with slogans or graphics that include hate speech or discriminatory content. • Risk that users could input targeted query seeking products in Shopping with slogans or graphics that include discriminatory content and view such products in the first few pages of results. • Risk that third party web content linked in News results includes hate speech or discriminatory content or content that perpetuates hateful attitudes or harmful stereotypes or otherwise facilitates discrimination towards individuals or groups. • Risk that Copilot Search results or suggested prompts reflect, reinforce, or perpetuate stereotypes, biases, or inequalities in the content generated • Risk that users leverage Bing Image or Video Creator or Copilot Search create discriminatory or hateful content • Risk that Bing Image or Video Creator erases cultural identities through homogenized content creation • Risk that Bing Image or Video Creator might not include or accurately represent diverse abilities of people, accelerating the creation of biased and discriminatory image-based content. • Risk that Bing Image or Video Creator perpetuates stereotypes or produces demeaning representations of individuals with autism or disabilities

	• Risk that Bing Image or video Creator might inadvertently portray genders (including trans, non-binary, etc.) in traditional or offensive roles based on biased training data. This risk can be increased if prompts not accurately or respectfully represent genders in non-English contexts. • Risk that Bing Image or Video Creator is used to produce images that portray certain races, genders, or religions in an unfair light if it's trained on biased data. • Risk that users leverage Bing Image or Video Creator to create discriminatory or hateful content • Risk that Bing Image or Video Creator is used to produce images that degrade or discriminate against specific groups. • Risk that GenAI summaries are based on low authoritative sources and produce content that may be harmful or misleading. • Risk that adversarial user queries or grounding data may bypass the safety features for Copilot Search and produce harmful or misleading content

Inherent Risk

The **Inherent Probability** of systemic risks related to Discrimination and Hate stemming from the use, misuse, or functioning of Bing's products and services absent sufficient mitigations is assessed as "Not Likely."

Bing considered internal metrics and social listening data as key factors in assessing a "Not Likely" level of probability relative to other potential systemic risks on Bing.

The **Inherent Severity** of content and activities that negatively impact Discrimination and Hate on Bing is rated as "High" primarily due to the gravity of risks within this risk area, considering the impact to wellbeing, societal, and economic impacts at the individual-level and societal impacts up to a regional scale. Risks within this risk area, such as hate speech, and discrimination may be remediable but can cause serious damage including loss of career advancement opportunities.

With the rating for Inherent Probability as "Not Likely" and Inherent Severity as "High," the **Inherent Risk** associated with Discrimination and Hate on the Bing service is "Moderate."

Maturity of Mitigations

Bing has implemented a robust set of mitigations to address risks related to Discrimination and Hate, including those described in Bing's Approach to User Safety, in Appendix II: Bing's Digital Safety Risk and Compliance Framework and Mitigations, and those risk-specific mitigations summarized below.

The mitigations Bing has implemented relative to Discrimination and Hate follow best practices with defined, documented, and managed processes. As such, the **Maturity of Mitigation** efforts Bing has applied to the Discrimination and Hate risk area is assessed as "Managed," which brings the **Residual Risk** rating down to "Low."

While not a proxy for Residual Risk, Bing has measured the DDR for Discrimination and Hate on Bing Search as an average of .14% from September 2024 to April 2025. This means that Bing estimates .14% of search traffic may include content that may have negative effects related to Discrimination and Hate.

Nevertheless, Bing continues to invest in developing and enhancing strategies to further mitigate and manage risks related to Discrimination and Hate. The highlights of the key mitigations currently implemented are described below.

Risk Specific Mitigations

The following risk-specific mitigations focus on measures uniquely tailored to address the identified risks in this area. These are in addition to the broader service-wide mitigations outlined in Bing's Approach to User Safety and the full inventory of Bing-wide protections listed in Appendix II. The mitigations described here include both unique interventions specific to Discrimination and Hate and expansions on previously described cross-platform mitigations with specific applications to Discrimination and Hate Where applicable, Bing's broader infrastructure has been adapted to address the particular risks associated with bias, hate speech, and discriminatory content. This layered approach ensures both platform-wide consistency and harm-specific responsiveness.

Bing's primary approach to reducing the visibility of Discrimination and Hate in search results focuses on promoting ***high-authority sources in search results***. As a part of Bing's commitment to a safe online environment and goal to ensure that users can access high-quality and authoritative content online, Bing's ranking principles consider low-quality content to include sources that use offensive statements, derogatory language, or name-calling to be low-quality.

Bing has a ***Threat Intelligence*** team that is accountable for implementing algorithmic interventions and metrics monitoring and dedicated to identifying and remedying, via targeted algorithmic interventions, high impact issues in search results, such as hateful speech, harassment against protected groups, and other problematic content that could have negative impacts. Bing employs a dedicated expert team that is dedicated to deploying targeted algorithmic interventions to reduce the prevalence of hate speech, and other problematic content in search.

Both Bing and specialized cross-Microsoft teams that support broadly across the company regularly engage with ***external stakeholders*** in areas of key policy priorities, such as terrorist and violent extremist content, information integrity and misinformation, CSEAI, RAI and AI-specific risks, hate speech, children and technology, and copyright and trademark infringement, to ensure that its internal policies, practices, and standards are addressing key concerns third-party stakeholders.

External engagements specific to discrimination and hate, Microsoft has engaged groups, such as the Global Project Against Hate and Extremism Project, the Anti-Defamation League (ADL), on issues related to hate speech, with a focus on how content policy and enforcement mitigate the risk of harm in content and GenAI.

Bing provides user transparency and more detailed information on the main parameters it uses for ranking in its ***transparency reports***, specifically the How Bing Ranks Your Content section of the Webmaster Guidelines, specifically calling out that in measuring the "quality" of a website, "pages that call for violence, name-calling, offensive statements, or use derogatory language to make a point are generally considered low-quality.

The ***Terms and Conditions for Users,*** found within the Microsoft Service Agreement (MSA), as applied to Discrimination and Hate prohibit using Microsoft services to engage in activity that is harmful to others (e.g., communicating hate speech). The Bing Image Creator Terms of Use further prohibit use of Bing's GenAI features to create or share hateful content; or to create or share content intended to exploit the vulnerabilities of individuals or groups. For example, without limitation, age, disability, or social or economic situations.

Mental and Physical Wellbeing

Risks related to Mental and Physical Wellbeing include exposure to content pertaining to gender-based violence, harmful behavioral suggestions, and content that encourages self-harm like suicide. Bing applies

its comprehensive safety mitigations to the wide range of web content that poses risks of negatively impacting users' Mental and Physical Wellbeing. By prioritizing high authority content, Bing's ranking algorithms are designed to protect users from harmful web content, such as content that promotes suicide or self-harm, or content that advocates gender-based violence. GenAI features deploy additional mitigations to prevent harmful AI-generated outputs.

Considering the Inherent Probability, Inherent Severity, and the Maturity of Mitigations relevant to Mental and Physical Wellbeing, the Bing Risk Assessment team has assessed the Residual Risk related to Mental and Physical Wellbeing on the service to be "Low."

Risk Area	Inherent Probability	Inherent Severity	Inherent Risk Rating	Mitigation Maturity	Residual Risk Rating
Mental and Physical Wellbeing	Not Likely	Critical	Moderate	Defined	Low

Risk Definition

Risk Area	Mental and Physical Wellbeing
Risk Definition	Risk that content or activities with actual or foreseeable serious negative consequences to a person's physical and mental well-being or in relation to gender-based violence occur on the service.
Risk Scenario	• Gender-based violence • Individual threats of violence, cyberbullying, harassment, stalking, public order offences, fear, or provocation of violence, and controlling or coercive behavior • Encouraging or assisting suicide or attempted suicide or serious self-harm or self-injury, including eating disorders • Distressing content or activities negatively impacting mental health • Design features encouraging excessive use or creating deeper browsing, potentially leading to negative mental or physical impacts • Content or activities that otherwise negatively impact Mental and Physical Wellbeing
Theoretical Risk Manifestations absent Mitigations	• Risk that search results or auto-suggest features lead users to web results that include gender-based violence. • Risk that search engines return content that could be used to locate content or communities focused on bullying or harassing people in the top search results. • Risk that search results or autosuggest features might lead users to web results that include content that instructs on, promotes, normalizes, praises, or facilitates suicide or self-harm. • Risk that autosuggest features could prompt users to engage more deeply with content promoting or glorifying self-injury. • Risk that search engines return content that promotes unrealistic or unhealthy body standards that could negatively impact mental health, in the top search results. • Risk that search engines return potentially distressing health or other information that can have a negative impact on users' mental health. • Risk that search engines return content that includes graphically violent images or content related to domestic abuse in the top search results. • Risk that Shopping listings show products with logos or graphics that promote include gender based violence. • Risk that News results show content that content that instructs on, promotes, normalizes, praises, or facilitates promotes suicide, self-harm, or causing harm to others.

	• Risk that News results show distressing content or activities that can negatively impact users' mental health. • Risk that Maps locations and listings could lead to offline physical harms. • Risk that Travel ideas, itineraries, and/ or search results suggest dangerous routes, destinations, or activities that place individuals in physical danger or otherwise undermine their physical well-being. • Risk that Advertisements appear on Bing that promote show addictive or harmful products or illegal substances that can negatively impact mental or physical wellbeing. • Risk that Bing Image or Video Creator is used to create content promoting gender-based violence or threatening violence • Risk that Bing Image or Video Creator is used to create material encouraging or depicting self-harm or suicide • Risk that Search or Bing Image or Video Creator results promote unrealistic or unhealthy body standards that could negatively impact mental health

Inherent Risk

The Inherent Probability of systemic risks related to Mental and Physical Wellbeing stemming from the use, misuse, or functioning of Bing's products and services absent sufficient mitigations is assessed as "Not Likely."

Bing considered internal metrics and social listening data as key factors in assessing a "Not Likely" level of probability relative to other potential systemic risks on Bing.

The **Inherent Severity** of content and activities that negatively impact Mental and Physical Wellbeing is rated as "Critical" primarily due to the gravity of risks within this risk area, considering the potential for significant harm to wellbeing, economic, and societal systems at the individual and potentially broader level. The impact of some risks within this risk area, such as violence against women, are irremediable. Other risks, such as behavioral addictions, may be considered remediable but still have the potential to cause significant damage to an individual's mental health or threaten their physical wellbeing.

With the rating for Inherent Probability as "Not Likely" and Inherent Severity as "Critical," the **Inherent Risk** associated with negative impacts to Mental and Physical Wellbeing on the Bing service is "Moderate."

Maturity of Mitigations

Bing has implemented a set of mitigations to address risks related to Mental and Physical Wellbeing, including those described in Bing's Approach to User Safety, in Appendix II: Bing's Digital Safety Risk and Compliance Framework and Mitigations, and those risk-specific mitigations summarized below.

The mitigations Bing has implemented relative to Mental and Physical Wellbeing follow best practices with defined and documented processes. As such, the **Maturity of Mitigation** efforts Bing has applied to the Mental and Physical Wellbeing risk area is assessed as "Defined," which brings the **Residual Risk** rating down to "Low."

The Mitigation Maturity for Mental and Physical Wellbeing is rated as "Defined" due to the risks related to a new GenAI product, Bing Video Creator, which introduces potential risk due to its inherent newness and the ability of users to use GenAI products in potentially new and unforeseen ways despite the mitigations currently in place.

While not a proxy for Residual Risk, Bing has measured the DDR for Mental and Physical Wellbeing on Bing Search as an average of .07% from September 2024 to April 2025. This means that Bing estimates .07% of search traffic may include content that may have negative effects related to Mental and Physical Wellbeing.

Nevertheless, Bing continues to invest in developing and enhancing strategies to further mitigate and manage risks related to Mental and Physical Wellbeing. The highlights of the key mitigations currently implemented are described below.

Risk Specific Mitigations

The following section highlights both how broader service-wide mitigations described in Bing's Approach to User Safety and inventoried in Appendix II, apply for this risk area, and include mitigations that are uniquely tailored to address the identified risks in this area.

The mitigations described here include both unique interventions specific to Mental and Physical Wellbeing and expansions on previously described cross-platform mitigations with specific applications to Mental and Physical Wellbeing. Where applicable, Bing's broader infrastructure has been adapted to address the particular risks associated with gender-based violence, cyber-bullying, and self-harm. This layered approach ensures both platform-wide consistency and harm-specific responsiveness.

Bing designs its ranking and recommendation algorithms to align with core product principles that prioritize ***high-quality, relevant content***, and to ensure that users are not offended, harmed, or misled by problematic material in search results.

Bing adds information to help ensure that users are not harmed by materials returned to search queries and will ***point to authoritative information***, when search results lack high authority content, or redirect users when they search for specific queries. For example, when users search for information on suicide methods they are redirected to suicide prevention hotlines, which are available in 31 countries. This also applies to images or videos related to suicide, and search queries self-harm, self-harm images or videos. Queries related to self-harm or suicide are considered high-risk queries, and harmful content around these queries are considered low quality, and generally demoted but demoting user-generated content is tricky for image and videos due to quality control signals from social media but the PSA to redirect users to help lines are shown. Similarly, PSAs are also shown for user search queries potentially related to searching for Child Sexual Abuse Material, directing users to hotlines and report. For Copilot Search, queries related to self-harm or suicide are considered high-risk, and Copilot Search responds to these queries through "thoughtful disengagement" and points users to support options. For search queries regarding eating disorders, under Bing's content policies this falls under a form of self-harm, as it can be considered a physical risk for users, thus defensive search interventions are employed to direct users towards high authority content. For responses generated by Copilot Search, there is a high authority approach taken regarding queries relating to eating disorders, to explain the negative effects of certain associated behaviors.

Bing's established "***Right to Be Forgotten***" policies and procedures help reduce risks of negative effects to Mental and Physical Wellbeing by providing reporting tools for users to submit requests that certain online content associated with its name be removed from Bing.

For ***industry engagement***, Microsoft has developed tools and multistakeholder partnerships to combat the rise of misinformation that can pose risks to Public Health or negatively impact Mental or Physical Wellbeing. Partnering with independent third-party organizations empowers Bing product teams to take additional actions to promote more authoritative information.

The ***Terms and Conditions for Users,*** found within the Microsoft Service Agreement (MSA), as applied to Mental and Physical Wellbeing prohibit using Microsoft service to publicly display, generate, or share inappropriate content or material (involving, for example, self-harm) and to engage in activity that is harmful to you or others (e.g., stalking, trying to generate or sharing content that harasses, bullies or threatens others, or advocating violence against others). The Bing Image Creator Terms of Use further prohibit use of Bing's

GenAI features to engage in activity that is harmful to the user, or others; to create or share content that could be used to harass, bully, abuse, threaten, or intimidate other individuals, or otherwise cause harm to individuals, organizations, or society; or to create or share glorification of violence.

Private and Family Life

Risks related to Private and Family Life include exposure to content pertaining to the malicious sharing or exploitation of private data, doxing, disclosure of private images (NCII), and privacy intrusions.

Bing's principles allow for removal of third-party search results that have an undue impact on privacy, such as content that can be used for ID theft or fraud, sensitive content that was inadvertently disclosed such as emails or health records, and nonconsensual pornography (NCII). In the EU, Bing respects the Right to be Forgotten.

Risks associated with NCII have shifted in the last year, with increased concerns from the public and regulators about how GenAI exacerbates these risks. In addition to the NCII content, stakeholders are also concerned about access to sites and applications that enable the creation of NCII, so called nudify apps.

Bing made investments in the assessment period to better identify and manage NCII-related content. Bing has a formalized NCII policy that addresses both real and "deepfake" images and is applicable across its ecosystem. Bing also introduced interventions to improve detection and management of NCII, including partnering with StopNCII to detect and remove known NCII based on hashes. Bing also improved the defensive search interventions to better identify prompts and as appropriate, remove potentially harmful content related to NCII and reduce the visibility of nudify apps as low-quality content.

Bing's report a concern form was also updated to capture NCII separately from reporting other privacy issues. These efforts form part of an evolving framework incorporating human review, clear policy enforcement, and low-latency user appeals. Despite these advances, Bing continues to explore ways to enhance its ability to address the complex risks posed by NCII.

As mentioned, Bing has invested in layered defenses for its GenAI capabilities, including upstream filters to detect jailbreak attempts, midstream blocking of high-risk prompts, and downstream controls limiting generation of content related to public figures or harmful subjects. These targeted mitigations, combined with ongoing measurement pipelines and monitoring systems, are regularly refined to ensure effectiveness without compromising user access to information, as referenced above.

Private and Family Life's Residual Risk rating increased from the prior year assessment due to overall increase in the prevalence of NCII internally, as well as in discussions with regulators and civil society, and in online discourse.

Considering the Inherent Probability, Inherent Severity, and the Maturity of Mitigations relevant to Private and Family Life, the Bing Risk Assessment team has assessed the Residual Risk related to Private and Family Life on the service to be "Moderate" (See *Changes in Risk Rating* section for further details).

Risk Area	Inherent Probability	Inherent Severity	Inherent Risk Rating	Mitigation Maturity	Residual Risk Rating
Private and Family Life	Likely	High	High	Defined	Moderate

Risk Definition

Risk Area	Private and Family Life

MICROSOFT CONFIDENTIAL MSFT_JUD_000002515

Risk Definition	Risk that content or activities with actual or foreseeable negative effect on respect for Private and Family Life occur on the service.
Risk Scenarios	• Malicious sharing or exploitation of sensitive or private data and doxing • Disclosure or threats to disclose nonconsensual, private, or intimate imagery or information (NCII) • Privacy intrusions including intrusive marketing practices that exploit personal data • Content or activities that otherwise negatively impact Private and Family Life
Theoretical Risk Manifestations absent Mitigations	• Risk that search engines return content that is doxing individuals or could be used for malicious exploitation. • Risk that search results or autosuggest features lead users to content that includes NCII or leads users, including minors, to synthetic NCII content. • Risk that search engines return content that could be used to find applications, tools, or services for creating synthetic NCII in the top search results. • Risk that street-level imaging features in Maps may inadvertently compromise individuals' privacy. • Risk that practices of Advertisements on Bing, particularly those involving targeted or personalized ads, intrude upon individuals' private lives by making inferences or using or collecting data without explicit consent, potentially leading to unwelcomed privacy intrusions and undermining individuals' autonomy over their personal information. • Risk that private property is featured as a destination or things to do in Travel results. • Risk that bad actors misuse Copilot Search to accelerate the creation of content that could facilitate doxing, exposure of private or intimate information, or identity theft • Risk that Bing Image and Video Creator could be misused to generate synthetic, yet realistic private or intimate images of individuals • Risk that Bing Image and Video Creator could be misused to generate synthetic, yet realistic private or intimate images of individuals

Inherent Risk

The **Inherent Probability** of systemic risks related to Private and Family Life stemming from the use, misuse, or functioning of Bing's products and services absent sufficient mitigations is assessed as "Likely."

Bing considered internal metrics and social listening data as key factors in assessing a "Likely" level of probability relative to other potential systemic risks on Bing.

The **Inherent Severity** of content and activities that negatively impact Private and Family Life on Bing is rated as "High" primarily due to the gravity of risks within this risk area, considering the potential for harm to economic, security, societal, and wellbeing systems at the individual-level with some broader reverberations. Some risks in this risk area, such as posts that contain sensitive data or disclosure of nonconsensual private or intimate images, have potentially irremediable impacts. Some risks in this risk area, such as users' inadvertent sharing of sensitive personal information (e.g., General Data Protection Regulation ("GDPR"), Article 9 information that includes racial or ethnic origin, political opinions, religious or philosophical beliefs etc.) may be less severe.

With the rating for Inherent Probability as "Likely" and Inherent Severity as "High," the **Inherent Risk** associated with Private and Family Life on the Bing service is "High."

Maturity of Mitigations

Bing has implemented a robust set of mitigations to address risks related to Private and Family Life, including those described in Bing's Approach to User Safety, in Appendix II: Bing's Digital Safety Risk and Compliance Framework and Mitigations, and those risk-specific mitigations summarized below.

The mitigations Bing has implemented relative to Private and Family Life follow best practices with defined, documented, and managed processes. As such, the **Maturity of Mitigation** efforts Bing has applied to the Private and Family Life risk area is assessed as "Defined," which brings the **Residual Risk** rating down to "Moderate."

While not a proxy for Residual Risk, Bing has measured the DDR for Private and Family Life on Bing Search as an average of .02% from September 2024 to April 2025. This means that Bing estimates .02% of search traffic may include content that may have negative effects related to Private and Family Life.

Nevertheless, Bing continues to invest in developing and enhancing strategies to further mitigate and manage risks related to Private and Family Life. The highlights of the key mitigations currently implemented are described below.

Risk Specific Mitigations

The following section highlights both how broader service-wide mitigations described in Bing's Approach to User Safety and inventoried in Appendix II, apply for this risk area, and include mitigations that are uniquely tailored to address the identified risks in this area.

The mitigations described here include both unique interventions specific to Private and Family Life and expansions on previously described cross-platform mitigations with specific applications to Private and Family Life. Where applicable, Bing's broader infrastructure has been adapted to address the particular risks associated with NCII, doxing, and privacy intrusions. This layered approach ensures both platform-wide consistency and harm-specific responsiveness.

The ***Terms and Conditions for Users***, found within the Microsoft Service Agreement (MSA), as applied to Private and Family Life prohibit using Microsoft service to violate or infringe upon the rights of others (e.g., taking photographs or video/audio recordings of others without their consent for any other purpose using any of the Services) and engage in activity that violates the privacy of others. The Bing Image Creator Terms of Use further prohibit use of Bing's GenAI features to engage in activity that violates the privacy of others; to attempt to create or share content that could violate the privacy of others, including disclosure of private information (sometimes known as "doxing"); to conduct facial identification, or identification verification purposes; to input photographs or video/audio recordings of others taken without their consent for the processing of an individual's biometric identifiers or biometric information; or to engage in activity to infer a person's emotional state, age, race, political opinions, trade union membership, religious or philosophical beliefs, sex life, or sexual orientation.

Safety systems in Bing's GenAI features are designed to prevent abuse, by include measures such as blurring faces in images before they are used as prompts in visual search and preventing the generation of images with individual faces. Bing also labels images as generated by Bing Image and Video Creator to limit the possibility of misuse.

Specifically, before any new feature or product is introduced to users, Bing requires it to pass through a robust ***pre-launch risk assessment***. This assessment process is part of Microsoft's broader compliance framework and includes reviews of privacy, security, accessibility, safety, and responsible AI concerns.

These reviews are conducted by professional compliance managers who ensure that products meet Microsoft's standards and legal obligations. Where risks are identified, teams must implement appropriate mitigations before the product or feature can go live. This process ensures that safety, privacy, and integrity are designed into Bing experiences from the start.

For safety, ***Bing visual search features*** (which allow use of images as a search prompt) do not allow for matching of images of private individuals' faces with content in the search index. It also does not allow for searching of adult images. In the visual search feature, Bing blurs faces before processing data contained in files provided by the user.

Bing maintains rigorous and principled processes, policies, and procedures for ***content removal and other mechanisms to counter harms***. Central to these processes is a firm commitment to the prevention of CSEAI. Bing proactively scans content to detect known CSEAI using hash-matching technologies such as PhotoDNA and MD5. This scanning occurs before content enters the search index and is used to ensure such material does not appear in results at all, across Core Search and Copilot Search. Bing also proactively scans upload image uploads in Bing Image Creator and Bing Video Creator to prevent known CSEAI from being uploaded. This proactive detection is supplemented by intelligence from trusted partners, including NGOs and law enforcement, and through user reporting channels. When CSEAI is identified, Bing removes it globally and without hesitation. This is a zero-tolerance area where Microsoft's commitment is absolute.

Bing also works to remove adult or sexually explicit images of another person online when shared without that person's consent. This is commonly referred to as nonconsensual intimate imagery (NCII), or "revenge porn". Microsoft considers this to be a gross invasion of personal privacy – whether the content is real or a "deepfake" image that is created using AI or other photo editing tools

Finally, Bing also removes certain other content under global legal standards or Microsoft policies, such as materials containing extremely sensitive personal information that could be exploited for fraud or identity theft. Removals may be prompted by legal orders from governments, user-generated reports, or policy-based decisions following internal investigations. Microsoft applies rigorous standards to each request, including the potential impact on user rights and public interest.

Bing Image Creator uses the ***StopNCII initiative's*** database of hashes for reported and verified NCII content and continuously scans and removes NCII content from Bing's image index.

Bing's ***Report a Concern*** tool was updated during this period to include options for users to report non-consensual intimate imagery, in addition to the exposed personal or private information options previously included.

Bing engages in regular self-directed tests and audits of its system to ensure its ability to respond and Microsoft, including Bing, also responds to periodic evaluations by the third-party independent organization ***Ranking Digital Rights*** which publishes annual ratings on Bing's practices, governance, and leadership on the protection of freedom of expression and privacy.

Public Health

Risks related to Public Health include exposure to content containing legal but harmful substances and health misinformation. Content related to illegal substances is addressed in the Illegal Content and Activities section above. Bing applies its layered safety mitigations to content that poses risks to Public Health. Bing's ranking algorithms and GenAI safety mitigation systems are designed to protect users from harmful web content and prevent harmful or misleading GenAI outputs. Bing's threat monitoring and incident response processes are designed to continuously monitor evolving trends and place rapid interventions where needed.

Considering the Inherent Probability, Inherent Severity, and the Maturity of Mitigations relevant to Public Health, the Bing Risk Assessment team has assessed the Residual Risk related to Public Health on the service to be "Low," and furthermore did not see a change in the scoring for Public Health for this year.

Risk Area	Inherent Probability	Inherent Severity	Inherent Risk Rating	Mitigation Maturity	Residual Risk Rating
Public Health	Not Likely	Critical	Moderate	Managed	Low

Risk Definition

Risk Area	Public Health
Risk Definition	Risk that content or activities with actual or foreseeable negative effects on the protection of Public Health occur on the service.
Risk Scenarios	• Promotion of legal but harmful substances, products, or practices • False or misleading health information or misinformation • Health-related disinformation or content that maliciously undermines Public Health institutions or initiatives • Content or activities otherwise negatively impacting Public Health
Theoretical Risk Manifestations absent Mitigations	• Risk that search engines return content that can lead to physical harm, such as dangerous challenges in the top search results. • Risk that search engines return content that includes information about medical conditions, treatments, vaccines, or homeopathic remedies that are inaccurate, unsafe, or otherwise not appropriate or advisable based on an individual's condition in the top search results. • Risk that search engines return content that includes low quality information related to health conditions or medical advice in the top search results. • Risk that bad actors intentionally manipulate search results to increase the visibility of health misinformation, including outdated guidance on regulated goods. • Risk that Advertisements on Bing are used to promote legal but harmful substances. • Risk that Bing Shopping shows websites that include listings for counterfeit, gray market, or unlicensed pharmaceutical products, medical devices, supplements, or other products that could be detrimental to users' health • Risk that News shows content containing inaccurate or misleading information about public health issues, medical conditions, treatments, side effects, public health orders or decrees, or other information that could potentially undermine public health. • Risk that Bing Image or Video Creator is misused to generate misleading imagery related to public health issues • Risk that Bing Image or Video Creator results include information or images on legal but harmful substances or practices, or inaccurately represent the effects or consequences of harmful practices or substances • Risk that Bing Image or Video Creator is used to create images that glamorize the use of harmful but legal substances • Risk that GenAI features generate ungrounded results or mischaracterize cited web links that leads to misinformation about health. • Risk that GenAI features are based on low authoritative sources and produce content that may be harmful or misleading.

Inherent Risk

The **Inherent Probability** of systemic risks related to Public Health stemming from the use, misuse, or functioning of Bing's products and services absent sufficient mitigations is assessed as "Not Likely."

Bing considered internal metrics and social listening data as key factors in assessing a "Not Likely" level of probability relative to other potential systemic risks on Bing.

The assessment team concludes as time passes since the COVID-19 pandemic, public discourse related to Public Health risks has decreased on Bing's services, as indicated by social listening data, survey data, and internal metrics.

The **Inherent Severity** for content and activities that negatively impact Public Health on Bing is rated as "Critical," primarily due to the gravity of risks within this risk area, considering the potential for significant and irremediable harm to wellbeing, economic, and societal systems at the individual and up to global level. Some risks within this risk area, such as health misinformation and exposure to harmful substances, have potentially irremediable consequences.

With the rating for Inherent Probability as "Not Likely" and Inherent Severity as "Critical," the **Inherent Risk** associated with Public Health on the Bing service is "Moderate."

Maturity of Mitigations

Bing has implemented a robust set of mitigations to address risks related to Public Health, including those described in Bing's Approach to User Safety, in Appendix II: Bing's Digital Safety Risk and Compliance Framework and Mitigations, and those risk-specific mitigations summarized below.

The mitigations Bing has implemented relative to Public Health follow best practices with defined, documented, and managed processes. As such, the **Maturity of Mitigation** efforts Bing has applied to the Public Health risk area is assessed as "Managed," which brings the **Residual Risk** rating down to "Low."

While not a proxy for Residual Risk, Bing has measured the DDR for Public Health on Bing Search as an average of .04% from September 2024 to April 2025. This means that Bing estimates .04% of search traffic may include content that may have negative effects related to Public Health.

Nevertheless, Bing continues to invest in developing and enhancing strategies to further mitigate and manage risks related to Public Health. The highlights of the key mitigations currently implemented are described below.

Risk Specific Mitigations

The following section highlights both how broader service-wide mitigations described in Bing's Approach to User Safety and inventoried in Appendix II, apply for this risk area, and include mitigations that are uniquely tailored to address the identified risks in this area.

The mitigations described here include both unique interventions specific to Public Health and expansions on previously described cross-platform mitigations with specific applications to Public Health. Where applicable, Bing's broader infrastructure has been adapted to address the particular risks associated with health misinformation. This layered approach ensures both platform-wide consistency and harm-specific responsiveness.

Health responses are only generated using ***grounding data from trusted health domains*** which have been vetted by the Bing Health team.

Bing has a ***Threat Intelligence*** team that is accountable for implementing algorithmic interventions and metrics monitoring and dedicated to identifying and remedying, via targeted algorithmic interventions, high impact issues in search results, such as Public Health, and other problematic content that could have negative impacts.

For ***authoritative content and crisis hubs***, Bing actively directs users to trusted sources through several tools. For high-impact events such as various EU elections, Bing highlights answers and PSAs that point to expert-vetted information. During events like Public Health crises, Bing would apply defensive search

interventions such as demoting low authority results in ranking on queries that may lead users to data voids and misleading content informed by threat intelligence.

For ***industry engagement***, Microsoft has developed tools and multistakeholder partnerships to combat the rise of misinformation that can pose risks to Public Health or negatively impact physical or mental wellbeing. Partnering with independent third-party organizations empowers Bing product teams to take additional actions to promote more authoritative information.

The ***Terms and Conditions for Users***, found within the Bing Image Creator Terms of Use, as applied to Public Health, prohibit use of the service to engage in activity that is false, or misleading; to create or share content that could mislead or deceive others, including for example creation of disinformation; or to attempt to mislead others about whether the Creations were generated by the service.

Public Security

Risks related to Public Security include risks related to terrorism and violent extremisms as well as coordination of harm and misinformation related to crisis events. Bing applies its comprehensive, multi-layered safety mitigations to content that poses risks to public safety. Bing's ranking algorithms and GenAI safety mitigation systems are designed to protect users from harmful web content and prevent harmful or misleading GenAI outputs. Bing's threat monitoring and incident response processes are designed to continuously monitor evolving trends and place rapid interventions where needed. Additionally, Bing's partnerships with organizations like the Institute for Strategic Dialogue ("ISD") and Microsoft's leadership role in the Global Internet Forum to Counter Terrorism ("GIFCT") underscore its dedication to combatting terrorist, violent, and extremist content. Through these initiatives, Microsoft aims to contribute to the broader fight against online extremism and misinformation, aligning with Bing's commitment under the COPD to adapt to the challenges posed by GenAI.

Considering the Inherent Probability, Inherent Severity, and the Maturity of Mitigations relevant to Public Security, the Bing Risk Assessment team has assessed the Residual Risk related to Public Security on the service to be "Low;" which, as mentioned above in the *Changes in Risk Rating* section, experienced a Residual Risk Rating decrease from a "Moderate" level.

Risk Area	Inherent Probability	Inherent Severity	Inherent Risk Rating	Mitigation Maturity	Residual Risk Rating
Public Security	Not Likely	Critical	Moderate	Managed	Low

Risk Definition

Risk Area	Public Security
Risk Definition	Risk that content or activities with actual or foreseeable negative effects on Public Security occur on the service.
Risk Scenario	• Content or conduct that promotes, encourages, or facilitates violent extremist or terrorist activities; violent extremist or terrorist recruitment, funding, training, or incitement; violent extremist or terrorist imagery or content; or cyberterrorism. • Organized crime, cyberattacks, coordination or cooperation to inflict physical harm or death, or compromise public infrastructure and security • Misinformation related to crisis events • Disinformation undermining public security institutions or initiatives or related to crisis events • Content or activities that otherwise negatively impact public security

Theoretical Risk Manifestations absent Mitigations	• Risk that search engines return terrorist recruitment content. • Risk that search engines return terrorist- and violent extremist-operated websites that disseminate propaganda or recruit members. • Risk that search engines return content that promotes violent acts. • Risk that search engines return content that includes information on enabling users to build harmful weapons or devices for use in a terrorist or violent extremist attack. • Risk that search engines return content where online groups or communities engage in ideation and planning for a mass-violence attack. • Risk that Autosuggest features lead users to web results that instruct on how to work with others to avoid detection while planning violent attacks. • Risk that search engines return content that is inaccurate, misleading information, harmful misinformation, related to crisis events or issues of public security in the top search results. • Risk that search engines return content that includes disinformation related to crisis events or issues of public security in the top search results. • Risk that activities on Shopping are used to collect funds used to finance terrorist organizations. • Risk that Shopping is used to assemble harmful weapons or devices for use in a terrorist or violent extremist attack • Risk that Maps and Travel show locations or rental listings that are targeted for terrorist training, funding, or support for terrorist acts. • Risk that News results include inaccurate, misleading information or issues of general public security in EU member states and around the world. • Risk that News results include disinformation related to crisis events or issues of general public security • Risk that Maps' aerial and street level imaging of sensitive locations, including military bases, may expose national security assets. • Risk that Bing Image or Video Creator is used to generate images that contribute to public security risks • Risk that Bing Image or Video Creator is used to create Terrorist propaganda or imagery • Risk of Bing Image or Video Creator being used to create realistic threat scenarios or emergency event simulations that could cause public panic or hinder emergency response efforts • Risk that Bing Image or Video Creator is used to generate infographics or images that perpetuate a misrepresentation of public security events • Risk that Bing Image or Video Creator is used to generate visual propaganda that is used to disseminate mis- or disinformation related to crisis events • Risk that GenAI summaries generate ungrounded results or mischaracterize cited web links that leads to misinformation about public safety incidents. • Risk that GenAI summaries are based on low authoritative sources and produce content that may be harmful or misleading. • Risk that adversarial user queries or grounding data may bypass the safety features for Copilot Search and produce harmful or misleading content

Inherent Risk

The **Inherent Probability** of systemic risks related to Public Security stemming from the use, misuse, or functioning of Bing's products and services absent sufficient mitigations is assessed as "Not Likely."

Bing considered internal metrics and social listening data as key factors in assessing a “Not Likely” level of probability relative to other potential systemic risks on Bing.

The **Inherent Severity** of content and activities that negatively impact Public Security on Bing is rated as "Critical," primarily due to the gravity of risks within this risk area, considering the potential for significant harm to security, wellbeing, societal, political, and economic systems at the individual, country, and regional levels. The inherent impact of many risks within this risk area is irremediable.

With the rating for Inherent Probability as "Not Likely" and Inherent Severity as "Critical," the **Inherent Risk** associated with Public Security on the Bing service is "Moderate."

Maturity of Mitigations

Bing has implemented a robust set of mitigations to address risks related to Public Security, including those described in Bing's Approach to User Safety, in Appendix II: Bing's Digital Safety Risk and Compliance Framework and Mitigations, and those risk-specific mitigations summarized below.

The mitigations Bing has implemented relative to Public Security follow best practices with defined, documented, and managed processes. As such, the **Maturity of Mitigation** efforts Bing has applied to the Public Security risk area is assessed as "Managed," which brings the **Residual Risk** rating down to "Low."

While not a proxy for Residual Risk, Bing has measured the DDR for Public Security on Bing Search as an average of .07% from September 2024 to April 2025. This means that Bing estimates .07% of search traffic may include content that may have negative effects related to Public Security.

Nevertheless, Bing continues to invest in developing and enhancing strategies to further mitigate and manage risks related to Public Security. The highlights of the key mitigations currently implemented are described below.

Risk Specific Mitigations

The following section highlights both how broader service-wide mitigations described in Bing's Approach to User Safety and inventoried in Appendix II, apply for this risk area, and include mitigations that are uniquely tailored to address the identified risks in this area.

The mitigations described here include both unique interventions specific to Public Security and expansions on previously described cross-platform mitigations with specific applications to Public Security. Where applicable, Bing's broader infrastructure has been adapted to address the particular risks associated with terrorism and crisis events. This layered approach ensures both platform-wide consistency and harm-specific responsiveness.

Bing takes a multi-stakeholder approach by having ***External Partnerships related to Public Security***. For example, as a member of GIFCT, an initiative to bring together technology companies, governments, and experts to share means of countering terrorist and violent extremist from exploiting digital services, Bing has access to the GIFCT's Incident Response processes. This allows Bing to quickly become aware of, assess, and address potential content circulating online resulting from a terrorist or violent extremist event.

Bing participates in the Advisory Network (a multistakeholder advisory group) of the Freedom Online Coalition, a coalition of 37 governments working to advance Internet freedom.

Bing also works closely with the ISD to understand risks related to Terrorist, Violent, and Extremist Content ("TVEC") in search, as well as to surface "counter narratives" via ad grants that help deter users who indicate an intent to learn more about extremist organizations.

The ***Terms and Conditions for Users***, found within the Microsoft Service Agreement (MSA), as applied to Public Security, prohibit using Microsoft service to engage in activity that is harmful to others (e.g., posting terrorist or violent extremist content). The Bing Image Creator Terms of Use further prohibit use of Bing's

GenAI features to create or share terrorism and violent extremist content; to engage in activity that is false or misleading; to attempt to create or share content that could mislead or deceive others, including for example creation of disinformation; or to attempt to mislead others about whether the Creations were generated by the service.

Rights and Protection of Minors

Risks related to Rights and Protection of Minors include exposure of minors to harmful behavioral activities and content and collection of minors' data without parental consent. Bing is committed to prioritizing the rights and safety of minors, aligning its product designs with the principles of the United Nations Convention on the Rights of the Child ("UNCRC"). To protect young users from harmful web content, Bing offers a SafeSearch feature to enable filtering out adult and graphic or violent content. SafeSearch is available without authentication. Additionally, Microsoft Family Safety features enables parents to manage and track use of applications, including Bing.

Bing does not require users to authenticate to access search. This is standard practice across search engines and is of the utmost importance given the role that search engines play in enabling users to freely access information. Allowing users to search without authentication is also an important privacy consideration for both adults and minors as authentication requires users to submit personal information. Bing Search users who are not identified through authentication as minors are treated as adults.

Microsoft believes there is value in minors engaging with GenAI services. At the same time, Bing also understands that, without appropriate safeguards, minors may engage with these tools in harmful or undesirable ways. In the interest of the Rights and Protection of Minors, creative GenAI services for Bing Image and Video Creator features are available only to authenticated users over the age of 13 (or over the age of parental consent if higher for their jurisdiction). Full functionality for Bing Image and Video Creator is disabled for all unauthenticated users.

Copilot Search defaults to Core Search results for any complex or sensitive queries, which already has its own layered restrictions, and therefore does not require the same age-based limitations as Bing Image and Video Creator. As a result, Copilot Search is available to all signed-in users without explicit age restriction.

Considering the Inherent Probability, Inherent Severity, and the Maturity of Mitigations relevant to Rights and Protection of Minors, the Bing Risk Assessment team has assessed the Residual Risk related to Rights and Protection of Minors on the service to be "Low."

Risk Area	Inherent Probability	Inherent Severity	Inherent Risk Rating	Mitigation Maturity	Residual Risk Rating
Rights and Protections of Minors	Not Likely	Critical	Moderate	Defined	Low

Risk Definition

Risk Area	Rights and Protection of Minors
Risk Definition	Risk that content or activities with actual or foreseeable negative effects on the protection of minors or respect for the rights of the child occur on the service.
Risk Scenario	• Child sexual exploitation and abuse, grooming, and trafficking of minors • Collection of children's data without parental consent, or targeting children with personalized advertisements

	• Exposure of children to harmful content or activities, including gambling, exploitation, crime, graphic violence, human gore, terrorism, or violent extremism • Discrimination, cyberbullying, and harassment of minors • Promotion of self-harm or self-injury for minors • Unnecessary restriction from access to information • Content or activities that otherwise negatively impact the rights or protection of minors
Theoretical Risk Manifestations absent Mitigations	• Risk that search engines return content that instructs on, normalizes, promotes, or facilitates sexual exploitation, grooming, and abuse of a child or pedophilia. • Risk that algorithmic or ranking systems inadvertently elevate results that aid human traffickers in recruiting by promoting false job postings that appeal to minors such as modeling or acting. • Risk that search engines return content that contains minors' personal data which can be used by adult users to locate, bond, and meet with children or used by abusers to target and groom minors. • Risk that Bing's data processes harm minors or collect its data without parental consent. • Risk that search results or autosuggest features lead minor users to websites that normalize harmful third-party content, such as hate speech, extremist content, violent or graphic material, adult content, information promoting or soliciting illegal and regulated goods, self-harm or content inciting violence against others. • Risk that search engines return content that shows risky behaviors to minors, such as dangerous challenges in the top search results. • Risk that auto-suggested searches expose minors to content that perpetuates discrimination and cyberbullying. • Risk that search engines return content to minors that normalizes eating disorders in the top search results. • Risk that Bing's safety systems for search unduly limit minors' access to information. • Risk that Advertisements on Bing expose minors to harmful activities targeted for child sexual exploitation and abuse • Risk that Shopping, Maps, News, and Ads may harm minor users by collecting minor's data without parental consent and target children minors with personalized content. • Risk that Advertisements on Bing target minors. • Risk that Shopping, News, and Ads results expose minors to harmful third party content such as hate speech, extremist content, violent or graphic material, adult content, information promoting or soliciting Illegal and regulated goods, including: Weapons and firearms, recreational drugs and paraphernalia, prescription drugs and pharmaceuticals, gambling, alcohol and tobacco products. • Risk that those age-inappropriate activities are recommended for children and/or families in Travel guides, or on the Maps. • Risk that Shopping or News could expose minors to products or content that perpetuate discrimination. • Risk that Bing's safety systems for Ancillary Features unduly limit minors' access to information. • Risk that images of children which do not conceal their privacy are shown on the "Travel the World Virtually" or in the Maps features. • Risk that Bing Image or Video Creator generates harmful content inappropriate of/for minors • Risk that Bing Image or Video Creator minor user prompts could be stored or collected without parental consent. • Risk that Bing Image or Video Creator is used to create imagery that glorifies and promotes self-harm for minors. • Risk that Bing Image or Video Creator promotes risky behaviors among minors, such as dangerous challenges or trends, through compelling images

Inherent Risk

The **Inherent Probability** of systemic risks related to the Rights and Protection of Minors stemming from the use, misuse, or functioning of Bing's products and services absent sufficient mitigations is assessed as "Not Likely."

Bing considered internal metrics and social listening data as key factors in assessing a "Not Likely" level of probability relative to other potential systemic risks on Bing.

It is important to note that the Bing service is not known to be used by a large number of minors (roughly 3% of EU Bing MAU are known to be users under 18), and minors under the legal age of consent in their respective location are prevented from accessing the full slate of Bing GenAI features.

The **Inherent Severity** for content and activities that negatively impact the Rights and Protection of Minors on Bing is rated as "Critical" due to the gravity of risk in this risk area, the vulnerable nature of the population impacted, the irremediability of impact of certain risks, and the potential for societal impact broader than the individual scale for risks impacting youth. Not all risks within this risk area are assessed at the highest level of severity, namely collection, processing, and use of minors' data without parental consent.

With the rating for Inherent Probability as "Not Likely" and Inherent Severity as "Critical," the **Inherent Risk** associated with negative impacts to the Rights and Protection of Minors on the Bing service is "Moderate."

Maturity of Mitigations

Bing has implemented a set of mitigations to address risks related to the Rights and Protection of Minors, including those described in Bing's Approach to User Safety, in Appendix II: Bing's Digital Safety Risk and Compliance Framework and Mitigations, and those risk-specific mitigations summarized below.

The mitigations Bing has implemented relative to the Rights and Protection of Minors follow best practices with defined and documented processes. As such, the **Maturity of Mitigation** efforts Bing has applied to the Rights and Protection of Minors risk area is assessed as "Defined," which brings the **Residual Risk** rating down to "Low."

The Mitigation Maturity rating of "Defined" reflects the presence of numerous sophisticated, risk-specific mitigations. However, this rating also acknowledges certain limitations. One factor influencing the rating is that Copilot Search does not restrict minors from accessing the service, meaning Copilot Search services should be considered as part of the Rights and Protection of Minors Risk Area. Another factor is the recent removal of the risk of CSAM from the Rights and Protection of Minors risk area, with associated risks and mitigations now realigned under the Illegal Content risk area. These aspects, while not increasing residual risk materially, are taken into account in assessing the in scope mitigations.

While not a proxy for Residual Risk, Bing has measured the DDR for Rights and Protection of Minors on Bing Search as an average of .54% from September 2024 to April 2025. This means that Bing estimates .54% of search traffic may include content that may have negative effects related to Rights and Protection of Minors.

Nevertheless, Bing continues to invest in developing and enhancing strategies to further mitigate and manage risks related to Rights and Protection of Minors. The highlights of the key mitigations currently implemented are described below.

Risk Specific Mitigations

The following section highlights both how broader service-wide mitigations described in Bing's Approach to User Safety and inventoried in Appendix II, apply for this risk area, and include mitigations that are uniquely tailored to address the identified risks in this area.

The mitigations described here include both unique interventions specific to the Rights and Protection of Minors and expansions on previously described cross-platform mitigations with specific applications to the Rights and Protection of Minors. Where applicable, Bing's broader infrastructure has been adapted to address the particular risks associated with gender-based violence, cyber-bullying, and self-harm. This layered approach ensures both platform-wide consistency and harm-specific responsiveness.

Like other risk areas, Bing ***utilizes targeted algorithmic interventions*** to remedy high impact issues in search results that could negatively impact minors and reviews the efficacy of the interventions.

In response to user search queries for CSAM, Bing displays ***in-product indicators*** and PSAs to help protect minors, similar to those regarding negative searches regarding suicide, stating that CSAM is illegal and depending on the query may not return search results; Bing also includes links to report these illegal materials or receive anonymous support, aimed at those who are victims of CSAM.

Bing has implemented a comprehensive suite of measures aimed explicitly at ***safeguarding the rights, privacy, safety, and security of minors*** across its service. These measures include setting age restrictions under the Microsoft Services Agreement ("MSA"), developing age-appropriate privacy communications, ensuring Bing features are accessible and understandable to teen users, as well as enforcing strict advertising policies through Microsoft Advertising. Additionally, Bing's commitment to privacy protection for minors is evident through practices such as minimizing data use and engaging with external stakeholders on children's safety and development. Through these efforts, Bing demonstrates its dedication to creating a secure and supportive online environment for young users, while continuously seeking to enhance its protective mechanisms in alignment with evolving standards and stakeholder feedback, including Microsoft policy teams, regulators, and civil society, to continue to iterate and improve on trust and safety in the Bing service.

Bing Image and Video Creator is available to authenticated users over the age of 13 (or higher in jurisdictions with a higher age of parental consent). Any user signed in with a Microsoft account that identifies the user as under 13 years of age (or older, in jurisdictions that require parental consent for older teens) cannot access these services in an authenticated state, even in the limited version available to unauthenticated users.

Microsoft ***does not deliver personalized advertising to minors*** whose birthdate in their Microsoft account identifies them as under 18, in accordance with the Microsoft's Privacy Statement. Across Bing features, authenticated users under 18 are not subject to targeted advertising. This important protection for youth will be carried forward into the new Bing features as well. Per the Microsoft Advertising Remarketing policy, advertisers must also not (i) "create a remarketing list, retarget or otherwise profile" any individuals under the age of 18; (ii) target individuals under the minimum age required for the product advertised; and (iii) implement remarketing on sites or applications directed to minors under the age of 18 or other applicable age limitations in an applicable market. Similarly, Microsoft has clear policies in place to prevent showing ads that are behaviorally targeted or contain adult materials such as alcohol or gambling to known child users. Microsoft uses both manual and systematic methods to review ads for compliance with its policies and reserves the right to remove those that are non-compliant.

For ***pre-launch risk assessments***, regarding minors there are specific standards that must be adhered to. For example, privacy reviews ensure that minor users are defaulted to the highest level of privacy protection and that disclosures and consent experiences are written in child-friendly language.

The ***DPIA*** process also requires a review as to whether the service has appropriately addressed possible harms to minor users. Although, to Bing's knowledge, there are relatively few minors on the Bing service, Bing still works to comply with applicable Microsoft policies and data use requirements as to minor users, and to consider the best interests of the child in product development. For known child users, Bing sets default controls to the highest level of privacy protection. Additionally, users are not permitted to search for adult content using the visual search feature, which helps limit the possibility that users could find CSEAI on the service.

Bing's established "***Right to Be Forgotten***" policies and procedures help reduce risks to the rights of the child or minors by providing reporting tools for users to submit requests that certain online content associated with their name be removed from Bing.

Bing has ***externally engaged*** with other government agencies on issues pertaining to child safety, consumer protection, CSEAI, and NCII and these engagements are important tools to inform the design and performance of Bing safety systems.

Microsoft routinely engages with ***external experts on child online safety***, through bilateral conversations and in forums such as the Internet Governance Forum ("IGF"). For instance, Microsoft is a founding and active member of the Tech Coalition, focused on facilitating industry cooperation to address online child sexual exploitation and abuse, thus demonstrating a strong commitment to collaborative efforts in combating evolving CSEA challenges, where Microsoft routinely receive insights and information from other industry participants, as well as briefings from researchers.

Additionally, Microsoft facilitated an internal meeting with civil society partners, to explore two key topics,1) state privacy legislation, and 2) minor data and AI. Microsoft presented some draft guiding principles and foundational capabilities for developing AI products and services for young people. Feedback from external engagements has provided valuable insights into enhancing the user experience for young audiences. These include suggestions for improving geolocation accuracy and the readability of responses, furthering Bing's commitment to transparency and reliability in its digital offerings.

Microsoft also offers a child-friendly version of its ***Microsoft Privacy Statement*** to better inform younger users of data practices. Microsoft has also developed a bespoke privacy page for young people that talks to Microsoft's data collection practices in age-appropriate language.

Bing's ***Content Moderation Ops team carries out training*** for content reviewers to ensure alignment on policy, labelling, and enforcement. For high-consequence harms, like child sexual exploitation and abuse, specialized teams receive additional focused training and wellness resources.

The ***Terms and Conditions for Users***, found within the Microsoft Service Agreement (MSA), as applied to the Right and Protection of Minors prohibit using Microsoft service to engage in any activity that exploits, harms, or threatens to harm children.

Conclusion

Conclusion of Assessment

After evaluating the purpose and design of the Bing services, the key systemic risks associated with the functionality and features available on Bing, and the safety systems and other risk mitigation processes implemented by Bing (as outlined in the *Introduction, Bing Overview and Risk Profile* and *Bing Risk Results* sections), the Bing Risk Assessment team has determined that Bing has implemented mitigations that are reasonable, proportionate, and effective to address the identified risks on the service.

Key considerations in the overall conclusions include:

The nature of Bing as a search engine and the critical role that search engines play in access to information critical to functioning within society

To uphold free expression and access to information, Bing works to avoid removing content from search results except in limited, narrow scenarios where legal demands or other important interests warrant removal (See the *Risk Specific Mitigation Section* for *Freedom of Expression and Information above)*. Bing has long established processes to ingest reports of concern from users and other stakeholders, including governments.

Where users are seeking low authority content that could be misleading or harmful, Bing works to include additional contextual information to search results. These interventions may include for example, answers, PSAs, site-level warnings, indicators of content provenance, and other high-quality information to supplement what is returned in the main search results.

The low occurrence of harms on the service

Internal estimates suggest 0.22% of queries entered on Bing are likely to lead users to unexpected problematic content. It also noted that, although Bing actions around 100 million pages of content removals every six months, the vast majority of these (>95%) are in response to copyright infringement claims and includes a small number of pages removed due to local legal obligations regarding materials that could pose risks to electoral processes, civic discourse, or public security. Given that Bing indexes hundreds of billions of webpages, this amount represents far less than 1% of the total webpages indexed. The assessment therefore determined that the probability of the identified risks occurring is low.

Reliance on Authority Principles

Absent a clear intent to access specific content, Bing assumes an intent to find high authority results and responds to search queries with the highest authority, relevant content, and grounds GenAI responses in the same principles.

The lack of user-generated content and user-to-user interaction on the service

As previously noted within the *Core Search Features* section, since Bing generally does not allow users the ability to post content or communicate directly with other users, this further limits Bing's risk exposure as to many other systemic risk areas common to online services, such as content virality, impersonation, or user rights regarding content removals and appeals. The risk of virality, impersonation, and user appeals is therefore low, and as such Bing has (appropriately) not dedicated significant resources to terms of use enforcement or internal complaint resolution processes. Bing has robust processes in place that will ensure that any new features that may implicate user content will be proactively reviewed, and additional risk mitigations implemented prior to launch.

The importance of allowing free and unauthenticated access to the Bing search engine and the very low number of authenticated users of the service

As an online search engine, Bing's primary objective is to offer access to information across the web by providing relevant and high-quality results in response to user queries, whether users are authenticated or not authenticated (See *User Base and Market* section above). Given the societal importance of search engines, Bing places a high priority on providing free and open access to Bing's search services. That means that a large majority of Bing users are not authenticated. This has implications on Bing's risk profile relative to the identification of minor users of the service as well as other user base demographics but helps promote user privacy, freedom of expression, and free access to information.

Collaboration with key stakeholders to inform risk measurement, mitigation, and broader safety considerations

As previously mentioned in the *Research and Partnership* and *Risk Specific Mitigation* sections, Bing works with cross-Microsoft teams to ensure compliance with internal policies and standards in a variety of high priority policy areas, such as privacy, digital safety, RAI, information integrity, intellectual property, diversity and inclusion, and accessibility. New product features in Bing are reviewed for compliance with each of these requirements before features are launched. Bing works with these cross-company teams of subject matter experts in order to identify new areas of concern for mitigation and to understand emerging legal requirements in the markets where it operates.

Data provided by users in Bing is handled in accordance with Microsoft privacy and security standards to ensure compliance with GDPR and other laws, and to avoid data breach.

The importance of continuing to assess and manage emerging threats, new risk vectors, and manifestations of harm in search

Bing and its internal Microsoft partners work with external researchers and subject matter experts (including data sharing relationships) to supplement its internal knowledge of key subject areas, to quickly identify new threats, and to obtain feedback on the efficacy of its processes.

As beforementioned, after evaluating the purpose and design of Bing's core search, GenAI, and ancillary search features, the key systemic risks associated with the functionality and features available on Bing, and the safety systems and other risk mitigation processes implemented by Bing, the Bing Risk Assessment team has determined that, while there are opportunities to continue to improve and mature Bing's Trust and Safety systems to address systemic risks more effectively, Bing's existing measures are sufficient and proportionate to mitigate key systemic risks on the service.

Progress on Enhanced Mitigations from Bing's Last Report

In the 2024 Systemic Risk Assessment report, Bing outlined six primary areas for enhancement of mitigations during the initial assessment period. Bing has made enhancements in each identified area as described below.

Governance and Organizational Alignment

To support durable systemic risk mitigation, Bing has reinforced its internal governance structures. This included formalizing safety governance across the Microsoft AI organization. Roles and responsibilities across Bing and associated product teams have been clarified and documented within the SRA Playbook to ensure accountability where systemic risk intersects multiple service areas. This structural clarity supports proactive ownership and streamlines escalation paths in the context of compliance obligations under Article 35(f).

Risk Detection and Monitoring Infrastructure

Bing continued to scale its investment in systemic risk detection through enhanced classifier development and risk mapping. Core efforts included enhanced alignment of internal classifiers to systemic risk areas, and classifier development to better track evolving threats. These classifiers have been integrated into the DDR pipeline and support ongoing monitoring. Complementing these efforts was the development of dashboards for risk monitoring, operational feedback loops, and compliance traceability.

Documentation and Incident Response

A critical focus has been placed on improving documentation standards across the systemic risk lifecycle. This included mapping how risks manifest in features, detailing mitigations in place, and measuring their effectiveness. A new team member was onboarded to centralize internal policy documentation. Incident response protocols were standardized. This work supports Articles 34 and 35 by ensuring procedural transparency and organizational learning and addresses current inconsistencies in documenting responses to content-related or operational incidents.

User Reporting and Notice-and-Action Mechanisms

Bing is refining and expanding its existing user reporting mechanisms to better capture nuanced harms and improve mapping between user-submitted concerns and specific product features or systemic risk areas. These updates build on existing notice systems and aim to enhance usability, support more detailed internal risk analysis, and provide richer insight into the user experience of risk. While implementation depends on migration to Microsoft's DigiTS RaC system, the intent is to support richer insights into user experience of risk. These changes do not replace or imply the inadequacy of prior systems but serve to evolve and strengthen reporting capabilities over time.

External Engagement and Consultation

Consistent with Recital 90 Bing is continuing to engage with civil society groups and issue-area specialists with the goal to integrate diverse perspectives into risk assessment and mitigation design, especially in high-impact areas such as youth safety, misinformation, or hate speech. A tracking process is in place to capture expert input and ensure it informs the refinement of risk mitigation measures. This supports more evidence-based and inclusive risk management practices.

Feature-Level Risk Quantification and Crisis Response

Bing is building infrastructure to calculate MAUs segmented by key features, login state, and age group. This supports a more precise assessment of probability and potential impact for systemic risks as required under

Article 34. Separately, crisis response capabilities are being strengthened, particularly for Maps. For example, during active crises, location-based recommendations will be suppressed in the Local Guide carousel to prevent misinformation or safety risks. These targeted interventions ensure that Bing can respond swiftly and appropriately in emergent situations.

Looking Ahead to the Coming Year

Bing has made investments in digital safety and risk mitigation throughout this reporting period, and it looks forward to continuing to build upon and improve its approach to mitigating systemic risks in the EU and adapting to new challenges ahead.

The following areas have been identified for heightened focus for the next assessment period, in addition to standard risk mitigation monitoring and refinement processes. Bing is committed to addressing each of the topics identified below by establishing a working group to 1) further define and evaluate the associated risks, 2) explore enhanced mitigation options, 3) review mitigation options with internal and external subject matter experts as appropriate 4) develop action plans as needed to address focus areas, and 5) track implementation of the developed action plans.

Enhancing governance and pre-launch review.

With the rapid pace of experimentation, learning, and innovation in search, including Bing, the team continues to leverage existing internal governance structures, while also improving those to increase scale, agility, and efficiency, and to more easily demonstrate compliance with regulation and corporate policies.

Developing and refining policies to align user experience with Bing's Trustworthy Search Principles.

To further refine Bing's Trustworthy Search Principles as outlined in *Bing's Values and Commitments* section, Bing has developed a policy framework and process that enables more cohesion across the different experiences that could result from users' queries. Leveraging this framework, Bing is also engaging earlier and more predictably with teams on product changes. In addition, Bing continues to build out policy documentation and resources to more closely align with operations, creating a flywheel of continuous learning and improvement in its management of systemic risks.

Continuing investments in targeted interventions to better manage systemic risks.

Bing continually invests in refining and enhancing its limited, targeted interventions designed to manage systemic risks effectively. For example, Bing is focused on improving threat intelligence capabilities to better detect and anticipate emerging risks. Bing is also consistently refining query and result classifiers to improve accuracy in identifying potentially harmful results and to ensure safer, more reliable information surfaces.

Refining documentation, resources, and metrics to enable greater data-driven improvement.

Bing continues to evaluate and, as appropriate, update user-facing resources, such as reporting mechanisms, as well as internal metrics to enable more nuanced and context-aware data analysis. This increased granularity in data enables Bing to identify opportunities to improve the management of systemic risks.

Improving children's safety.

As a part of Bing's ongoing commitment to user safety, and bolstered by regulatory changes, Bing is updating the protections provided for users that are in Strict or Moderate mode in SafeSearch in Bing. We believe these changes will create a better experience for those users, mitigating the risk that they might unexpectedly

encounter harmful, offensive, or illegal content, while not limiting users' ability to see or access sensitive content with SafeSearch in Off or Moderate.

Risk assessment and mitigation is and should be a continuously evolving process as threats and risks evolve and as Bing continues to engage in innovation and experimentation with respect to Search. Bing is committed to continuously monitoring the effectiveness of the implemented risk mitigation to address issues as they arise and continue to refine mitigations.

Appendix I: Detailed Systemic Risk Assessment Methodology

Risk Areas

Bing considered the following twelve risk areas for this Systemic Risk Assessment. In addition to risk definitions, Bing also identified common risk scenarios for each risk area to ensure adequate consideration of more specific risk or harm types within each risk area. Not all considered risk scenarios are applicable to each Bing product or service; however, they are used to explore the circumstances under which each risk outlined in Article 34 of the DSA could manifest on the service. As part of the risk assessment process, the Bing Risk Assessment team further examined whether and how each risk scenario could present on specific features of the Bing service to ensure adequate coverage of potential Inherent Risks with the implemented mitigations.

#	Risk Area	Risk Definition
1	**Civic Discourse and Electoral Processes**	Risk that content or activities with actual or foreseeable negative effects on Civic Discourse and Electoral Processes occur on the service.
2	**Consumer Protection and Fraud**	Risk that content or activities with actual or foreseeable negative impact to a high-level of consumer protection occur on the service.
3	**Discrimination and Hate**	Risk that content or activities with actual or foreseeable negative effect on the fundamental right to non-discrimination occur on the service, impacting individuals and groups based on race, ethnicity, religion, gender, sexual orientation, or other protected traits, or reflecting, reinforcing, or perpetuating harmful stereotypes, biases, or inequalities.
4	**Freedom of Expression and Information**	Risk that content or activities with actual or foreseeable negative effects on the fundamental right to Freedom of Expression and Information, including the freedom and pluralism of the media occur on the service.
5	**Human Dignity**	Risk that content or activities with actual or foreseeable negative effects on the fundamental rights to Human Dignity occur on the service.
6	**Illegal Content and Activities**	Risk related to the conduct of illegal activity or dissemination of illegal content through the service.
7	**Mental and Physical Wellbeing**	Risk that content or activities with actual or foreseeable serious negative consequences to a person's Mental and Physical Wellbeing or in relation to gender-based violence occur on the service.
8	**Private and Family Life**	Risk that content or activities with actual or foreseeable negative effect on respect for Private and Family Life occur on the service.
9	**Protection of Personal Data**	Risk that content or activities with actual or foreseeable negative effects on the Protection of Personal Data occur on the service.
10	**Public Health**	Risk that content or activities with actual or foreseeable negative effects on the protection of Public Health occur on the service.
11	**Public Security**	Risk that content or activities with actual or foreseeable negative effects on Public Security occur on the service.
12	**Rights and Protection of Minors**	Risk that content or activities with actual or foreseeable negative effects on the protection of minors or respect for the rights of the child occur on the service.

Risk Factors and Influencers

The European Commission has identified a variety of factors that may influence the probability of risks occurring on the service as well as the potential impact of various features of the service on systemic risks. As part of its assessment, the Bing Risk Assessment team considered the impact of the following risk factors and influencers on the potential manifestation of the identified systemic risks on Bing features, the probability of the risk occurring on Bing's service, as well as the mitigations are implemented to address the systemic risks:

- The design of recommender systems and their impact on the prioritization of content that users see on the service based on the designated criteria.
- The design of GenAI features and their impact on the ability of users to accelerate the generation of potentially harmful content, as well as serve content that is potentially harmful or misleading.
- Content moderation systems - whether manual or automated - and their impact on whether, how much, and what types of content are available to users on the service. This factor can influence both the amount of harmful content that users are exposed to on the service, and it can impact user's freedom of information and pluralism.
- Terms and conditions and their enforcement and their impact on what content is available to users on the service.
- Systems for selecting and presenting advertisements and their impact on user exposure to content, user privacy and consumer protection.
- Data-related practices and their impact on consumer protection from fraud and personal data protection.
- Intentional manipulation of the service, including inauthentic use or automated exploitation, impact on consumer protection, prioritized content, and potential volume of harmful AI-generated content.
- Amplification and potentially rapid and wide dissemination of illegal or violative content via recommendation or AI-generated content and its impact on systemic risk.
- As appropriate, regional, and linguistic considerations, such as the strength of content moderation in EU languages or the effectiveness of mitigation measures in EU member states.

Risk Assessment Inputs

This section details the evidence used to support the risk assessment scoring and rationale. The Bing Risk Assessment team collected information from a variety of sources to inform the risk assessment and has compiled this standard listing of inputs to ensure a consistent and comprehensive review of information as part of the annual assessment. The Bing Risk Assessment team gathered and reviewed the following inputs to inform and substantiate ratings assigned in the Systemic Risk Assessment.

- **Stakeholder validation of mitigations:** The Bing Risk Assessment team deployed an information collection sheet with associated questions to each Bing product team to systematically identify changes to Bing products and services since August 2024 to be described in the 2025 Systemic Risk Assessment.
- **Authoritative sources**: The Bing Risk Assessment team reviewed publicly available sources considered reliable due to their expertise and reputation, including the European Commission, CERRE and other sources such as such as civil society and other non-governmental organizations, on the severity of systemic risks.

- **Policies and publications:** The Bing Risk Assessment team reviewed relevant internal policies and publications, including the content policies on the Digital Safety at Microsoft page, public practices, and other publications, including blogs, to identify additional policies and initiatives most relevant to the risk assessment.
- **Mitigation summaries:** Bing stakeholders developed brief summaries of Bing controls and mitigations specific to each risk area.
- **Internal consultations:** The Bing Risk Assessment team conducted in-depth workshop sessions with internal stakeholder groups to solicit more detailed information on the presentation of risks and unique mitigations relevant to various features across Bing products and services. These internal consultations included representatives from Product Teams, Bing Compliance teams, and Microsoft Defensive Intelligence Response and Transparency team.
- **External consultations:** Bing regularly engages with external stakeholders, including civil society organizations, to receive feedback on the service and to discuss best practices for addressing risk.
- **Internal metrics:** The Bing Risk Assessment team considered internal metrics and measurements to inform both assessment of prevalence as well as effectiveness of mitigations.
- **Transparency report metrics:** The Bing Risk Assessment team considered metrics reported through Transparency Reporting to inform the assessment of both the probability and effectiveness of mitigations.
- **Open-source data on public discourse:** The Bing Risk Assessment team reviewed a collection of social and digital media articles and conversations, including social listening reports, around the Bing service and each systemic risk area to identify trends in areas of public discourse and/or concern.
- **Product and feature material changes:** The Bing Risk Assessment team evaluated product and feature material changes that occurred over the past year.
- **DSA Risk and Control Matrix:** The Bing Risk Assessment team incorporated controls outlined in the DSA Risk and Control Matrix and potential mitigations where relevant to systemic risks.

Probability

The Bing Risk Assessment team conducted a **data-driven probability assessment** to evaluate the likelihood of certain events occurring on the service absent mitigations by analyzing relevant data, including public incident data, transparency report metrics, and internal metrics.

Probability considers factors such as the inherent vulnerability and demand for the risk to occur on the service.

Assessment of **vulnerability** considers whether the perpetration of the harm on the service requires both intent and sophistication or whether it can occur without one or either. For example, can the harm accidentally occur with an average user conducting web searches online or does the harm require hacking skills with malicious intent.

Assessment of **demand** considers the volume of attempts or instances where this harm might occur. The Bing Risk Assessment team considered internal and external data samples for a data-estimated understanding of demand. For external data, the Bing Risk Assessment team reviewed public discourse on social and digital media in the EU of the considered risks in relation to Bing products and services to identify risk areas with a higher relative level of public concern. For internal data, the Bing Risk Assessment team reviewed internal metrics related to content reported by users, flagged, or removed for identified risks to identify risk areas with a higher relative of occurrence or attempted perpetration on the service.

The Bing Risk Assessment team considered the vulnerability and relative demand of the risks and compared the resulting ranking or probability to feedback from internal and external expert stakeholders, incorporating insights related to public events (for example the increased occurrence of elections this year), as well as user counts for Bing products and services, to validate the assigned probability rating.

Each probability rating is assigned a score from 1 to 5, with higher scores indicating a higher likelihood of the occurrence of the event on the service absent mitigations.

Inputs that substantiate the probability assessment include:

- Open-source data on public discourse
- Transparency report metrics
- Internal metrics
- Internal consultations
- External consultations
- Stakeholder validation of mitigations

Figure 14: Probability Rating Scale

Description	Weight	Score	Rating
The risk event or circumstance is relatively certain to occur without considering mitigations in place	>=243	5	Expected
The risk event or circumstance is highly likely to occur without considering mitigations in place	193 – 243	4	Highly Likely
The risk event or circumstance is likely to occur without considering mitigations in place	143 – 193	3	Likely
The risk event or circumstance occurring is possible but not likely without considering mitigations in place	93 – 143	2	Not Likely
The risk event or circumstance is only remotely probable without considering mitigations in place	<93	1	Remote

Severity

The Bing Risk Assessment team conducted an **objective, systems-based analysis** to determine severity, considering the complexity, scale, and gravity of impact to assign an overall severity rating.

Severity is calculated once for each risk and considers factors of complexity, scale, and gravity. Complexity is determined by the number of systems impacted by the harm. The Bing Risk Assessment team considered impact on economic, security, political, societal, environmental, and wellbeing systems. Scale is determined by whether the harm takes place at the individual, local, country, regional, or global level. Gravity considers the potential irremediability of the harm.

Figure 15: Severity Scoring Sheet

	Impact Scale				
Systems Impacted	Global	Regional	Country	Local	Individual
Economic	Low	Low	Moderate	High	High

MICROSOFT CONFIDENTIAL MSFT_JUD_000002537

Security	Low	Moderate	High	High	Moderate
Societal	Low	Moderate	High	High	Moderate
Political	Low	Moderate	High	High	Moderate
Environmental	Low	Low	Low	Low	Low
Wellbeing	Low	Low	Low	Low	Moderate

The Severity Scoring Sheet standardizes the consideration of the severity factors. The weighted output of the Severity Scoring Sheet corresponds with a Severity Rating on a graduated scale to avoid under-rating risks that do not have a regional or global impact but still have significant impact on several systems.

The Severity Ratings correspond to a score from 1 to 5, with higher scores indicating a higher level of Inherent Severity.

Inputs that substantiate the severity assessment include:

- External consultations
- Internal consultations
- Authoritative sources

Figure 16: Severity Rating Scale

Description	Weight		Score	Rating
	Min	Max		
Impact that could cause critical, irremediable harm, damage, or loss	145	270	5	Critical
Impact that could cause significant irremediable harm, damage, or loss	95	145	4	High
Impact that could cause some harm or disruption but is generally manageable or remediable	75	95	3	Moderate
Impact that could cause limited harm or disruption	71	75	2	Low
Impact that has little or no consequence	70	71	1	Minimal

Inherent Risk

Inherent Risk is a multiplication of Inherent Probability by Inherent Severity to identify the level of risk absent mitigations.

Figure 17: Inherent Risk Rating Scale

Description	Score (>=)	Score (<=)	Rating
A risk that is relatively certain to occur, would have a severe impact if it occurred, and requires immediate action to manage or mitigate.	17	<=	Critical

A risk that is highly likely to occur, would have a significant impact if it occurred, and requires urgent action to manage or mitigate.	11	17	High
A risk that is likely to occur, would have a noticeable impact if it occurred, and requires some action to manage or mitigate.	6	11	Moderate
A risk that is not likely to occur, would have a limited impact if it occurred, and may require minimal action to manage or mitigate.	2	6	Low
A risk that is remotely probable, would have a minimal impact if it occurred, and may not require any action to manage or mitigate.	0	2	Minimal

Maturity of Mitigations

The Bing Risk Assessment team then assessed the maturity of Bing's controls, safeguards, and other measures to mitigate the identified Inherent Risks in a reasonable, proportionate, and effective manner.

To assess the maturity of Bing's controls, the Bing Risk Assessment team collected enterprise-level mitigations, risk-specific mitigations, and feature-specific mitigations currently implemented and aligned them to the categories of the Bing Risk and Compliance framework to ensure coverage across industry-standard mitigation practices. Utilizing the Bing Risk and Compliance Framework enables Bing to set standards and enact best practices from peers across the technology industry, enhanced privacy protocols, balanced public discourse, and robust user protection mechanisms. Its wealth of resources empowers Bing to mitigate harmful content or actions effectively without infringing upon the freedom of expression.

Bing stakeholders complete product specific information collection sheets to identify implemented best practices within the listed Bing Risk and Compliance Framework best practice areas. Bing does not only develop and implement mitigations aligned with this framework; rather, the framework serves as a useful assessment to identify areas in which Bing can continue to mature implemented mitigations.

Figure 18: Bing Risk and Compliance Framework Best Practices Screenshot

Governance & Oversight			Product/Service Design and Implementation				External Actions
Governance	Compliance Management	Risk Management	Responsible Product/ Service Design	Policies & Other Mechanisms to Counter Harms	Content & Activity Moderation	Data Governance	External Engagement
Governance & Oversight Convene governance forum, Senior Management oversight function	**Compliance Policies, Frameworks & Enablers** Define internal governance policies and operating models	**Risk Assessment & Identification** Conduct programmatic and systemic risk assessment(s), identify and prioritize risk	**Product / Service Development Lifecycle** Conduct safety and compliance reviews prior to product or service launch and throughout development lifecycle	**Policies & Policy Development** Develop policies for unacceptable service/product use and response to violative content/actions	**Policy Enforcement Mechanisms & Infrastructure** Establish activities to monitor and address violative content/activity with transparency	**Data Management Lifecyle** Define ownership, handling, retention, removal, and destruction protocols	**Public Reporting and Transparency** Publish transparency reports including ads repository, Systemic Risk Report, SOR Database
Roles and Responsibilities Determine single points of contact, legal rep, and management body / board oversight	**Communication & Training** Communicate and educate internal staff on roles, responsibilities, and requirements	**Risk Decisioning & Mitigation** Determine risk tolerance, decision and mitigation pathways, and escalation	**Product Features and Functions** Design products (algos, recommenders) with functional / transparency requirements in mind and limit platform misuse	**Terms & Conditions** Ensure application of fair, reasonable, and non-discriminatory general terms and conditions of access for users	**User Reporting** Enable and effectively address user reporting of violative content /activity	**Privacy, Access, and Acceptable Use** Adopt baseline controls on user access and data portability, enable third party access to data	**Public & Civil Society Engagement** Publish annual transparency reports, consult with civil society on systemic risk and impact, share data with researchers
Planning & Resourcing Determining adequate resourcing across the firm to meet obligations	**Compliance Assurance** Conduct controls testing, refer remediation to relevant teams	**Risk Monitoring & Reporting** Establish continuous monitoring and report on changes in risk status	**Responsible AI** Policies, procedures, processes and tools to help ensure responsible use, monitoring and maintenance of data and AI	**User Education and Empowerment** Inform and empower users on safe platform use, provide user choices and system optionality	**Appeals & Complaints** Establish complaint and redress mechanisms and out of court dispute settlement	**User Choice & Transparency** Adopt baseline controls that helps users build autonomy over how their data is used and processed	**Regulatory Response & Engagement** Satisfy external reporting requirements, respond to data access requests
Regulatory Convergence & Change Management Integrate upcoming requirements into existing compliance infrastructure	**Issues / Incident Management & Monitoring** Validate that program enhancements proceed timely; escalate to governance forums	**Crisis Response** Develop and implement protocols for addressing crisis situations					**External Audit** Facilitate external audit and refer remediation to 1 LOD product teams
Fees and Admin Requirements Pay applicable supervisory fees, purchase insurance and other admin requirements							**External Mitigation** Coordinate and share data with national authorities as appropriate

The Bing Risk Assessment team then mapped the mitigations to the identified risk scenarios to identify implemented mitigations to address identified Inherent Risks. And finally, the Bing Risk Assessment team evaluated the implemented mitigations according to the DTSP maturity rating scale. The strength of the mitigations implemented covering the identified risks and the industry best practices would correspond to a higher maturity rating; however, regardless of the strength of existing mitigations, any risk area missing mitigations across any best practice or for any identified risk scenario cannot receive a rating higher than Managed maturity.

Figure 19: DTSP Maturity Rating Scale screenshot

(1) Ad Hoc	(2) Repeatable	(3) Defined	(4) Managed	(5) Optimized
A rating of Ad Hoc is assigned when execution of best practices is incomplete, informal, or inconsistent.	A rating of Repeatable is assigned when execution of best practices occurs without standardized processes. Organizations aim to document more formalized practices.	A rating of Defined is assigned when execution of best practices occurs with defined and documented processes. Processes are more proactive than reactive and are implemented across the organization.	A rating of Managed is assigned when execution of best practices is defined, documented, and managed through regular reviews. Organizations use feedback to continuously mitigate process deficiencies.	A rating of Optimized is assigned when execution of best practices promotes Trust & Safety in every aspect. Processes are continuously improved with innovative ideas and technologies.

Each Maturity Rating corresponds to a percentage score, with a lower percentage indicating a lower level of maturity.

Inputs that substantiate the severity assessment include:

- Stakeholder validation of mitigations.
- Guidance from authoritative sources, including the European Commission, on the severity of systemic risks.
- Collection of Bing's published policies and official communications.
- Bing internal stakeholder summarized policies and procedures.
- Notes from internal consultations.
- Notes from external consultations.
- Bing internal metrics.
- Transparency Reporting metrics.
- Open-source data on public discourse related to Bing and systemic risk areas.
- Bing product and feature material changes.
- DSA risk and control matrix.

Figure 20: Mitigation Maturity Scale

Description	Risk Coverage	Framework Coverage	Final Score	Rating
A rating of Ad Hoc is assigned when execution of best practices is incomplete, informal, or inconsistent.	0	N/A	10%	Ad Hoc

A rating of Repeatable is assigned when execution of best practices occurs without standardized processes. Organizations aim to document more formalized practices.	5	N/A	30%	Repeatable
A rating of Defined is assigned when execution of best practices occurs with defined and documented processes. Processes are more proactive than reactive and are implemented across the organization. A risk can only receive a Defined coverage rating if there are mitigations in place on each product to address all risks relevant to that product.	10	N/A	50%	Defined
A rating of Managed is assigned when execution of best practices is defined, documented, and managed through regular reviews. Organizations use feedback to continuously mitigate process deficiencies. A risk can only receive a Managed rating if, in addition to having full risk coverage, the mitigations in place cover 50% of framework categories.	15	50%	65%	Managed
A rating of Optimized is assigned when execution of best practices promotes Trust & Safety in every aspect. Processes are continuously improved with innovative ideas and technologies. A risk can only receive an Optimized rating if, in addition to having full risk coverage, the mitigations in place cover 100% of framework categories.	N/A	100%	75%	Optimized

Residual Risk

Inherent Risk is then multiplied against the Maturity of Mitigations score to identify a Residual Risk Rating that will help Bing to prioritize and identify which risks may require enhanced mitigation efforts.

Figure 21: Residual Risk Rating Scale

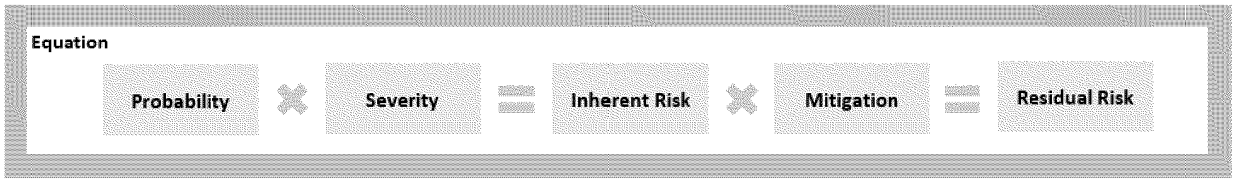
Description	Score (>)	Score (<=)	Score (<=)
A risk that is relatively certain to occur would have a severe impact if it occurred and requires immediate action to manage or mitigate.	15.1	<=	Critical
A risk that is highly likely to occur, would have a significant impact if it occurred, and requires urgent action to manage or mitigate.	10.1	15	High
A risk that is likely occur, would have a noticeable impact if it occurred, and requires some action to manage or mitigate.	5.1	10	Moderate

A risk that is not likely to occur, would have a limited impact if it occurred, and may require minimal action to manage or mitigate.	2	5.1	Low
A risk that is only remotely probable, would have a minimal impact if it occurred, and may not require any action to manage or mitigate.	0	2	Minimal

Sample scoring calculation

In this scoring calculation, Probability is multiplied by Severity to achieve Inherent Risk. Inherent Risk is multiplied by 1 minus Mitigation to achieve Residual Risk.

Figure 22: Systemic Risk Scoring Equation Screenshot



MICROSOFT CONFIDENTIAL MSFT_JUD_000002542

Appendix II: Bing's Digital Safety Risk and Compliance Framework and Mitigations

Below is a summary of Bing's mitigations that span multiple systemic risk areas. The mitigations below have been categorized across the Bing Safety Funnel. Although mitigations may address more than one component of the Bing Safety Funnel best practice, mitigations have been categorized based on the Bing Safety Funnel component it primarily addresses.

Product Quality

Terms and Conditions – Generative Experiences: GenAI Terms and Conditions governing the use of Bing's GenAI features are outlined in the Bing Image Creator Terms of Use. These provisions prohibit activities that could negatively impact users, such as engaging in harmful, fraudulent, or illegal activities, violating privacy, or creating inappropriate content. Users who violate these terms may face limitations on their access to the services.

Terms and Conditions – Microsoft Services Agreement: Bing publishes and enforces its content standards through terms and conditions, specifically through the Microsoft Services Agreement and the Bing Image Creator Terms of Use which, among other conditions, prohibit the generation or sharing of inappropriate or harmful content. These policies apply globally and are available in local languages. Bing reserves the right to suspend or ban users who attempt to misuse the service. For example, entering prompts designed to bypass safety systems is not permitted. These terms help ensure that Bing remains a safe and transparent service for all users.

Content and Activity Moderation:

Safety Measures – Search Suggestions: Bing has implemented measures aimed at curbing the generation of search suggestions that might tarnish users' fundamental rights. This includes creating avenues for users to effortlessly report suggestions they deem inappropriate, helping to ensure such concerns are promptly addressed. Through these actions, Bing demonstrates its commitment to fostering a respectful digital space where users feel valued and protected.

SafeSearch: Bing also offers users direct control over what they see through SafeSearch, a feature that uses machine learning to automatically detect and filter adult and explicit content from search results. This feature is especially important in educational and household settings where additional safeguards may be required.

SafeSearch settings can be locked to prevent tampering and ensure consistent application of filters. Bing also provides a mechanism for users to report inappropriate content or suggestions. These reports feed into the continuous improvement of the underlying safety systems, helping Bing respond quickly to emerging issues.

Content Provenance & Watermarks: Bing invests in helping users identify AI-generated content. Bing includes content credentials and a pixel-based watermark on generated images. The hidden watermark is imperceptible to human eyes, allows verification of AI generated images even after minor modifications. The Watermark complements Bing's content provenance technology to help ensure an AI-generated image's integrity and traceability. Bing participates in cross-industry efforts to improve identification techniques like the C2PA.

Metaprompting: To guide GenAI models in generating responses that align with Microsoft's AI Principles and user expectations, Bing employs a technique known as "metaprompting." Metaprompting involves providing instructions to the AI model to influence its behavior, ensuring that the system operates responsibly and ethically.

Safety by Design – GenAI Features: The development and deployment of GenAI features at Bing are guided by safety processes that ensure adherence to high standards of responsibility and emphasize safety by design for GenAI features. These processes include red team testing, expanded reactive social listening systems, prominent reporting functionality, and operations and incident response mechanisms. These measures help identify vulnerabilities and inform the development of targeted mitigations.

Content Moderation – Real Estate: Bing actively scans rental listing descriptions and images to ensure compliance with its content policies, employing automated tools to detect inappropriate content, hate speech, or sudden images. Bing Real Estate automatically detects fraudulent content and continues to improve Real Estate's fraud detection capabilities as new methods for posting harmful content arise. This scanning is complemented by allowing users to report listings that they find inappropriate or fraudulent.

Data Governance:

GenAI Disclosures: Copilot Search and Bing Image and Video Creator provides in-service disclosures and reminders to users that AI can make mistakes and to double check the veracity of information generated by AI. In addition, the Copilot Search FAQs, help pages, and other public facing information sources (such as blog posts and marketing materials) help educate users on the nature of AI-driven search experiences and the uses, safeguards, and limitations of this emerging technology, regularly reminding users of the potential for mistakes and risks of over-reliance.

Citation Links: Copilot Search responses include citation links, and the corresponding web results are displayed at the bottom of the results. This practice allows users to verify the information and assess the credibility of the sources.

User Privacy Choices: Respecting user privacy choices is a core Microsoft commitment, and as part of Product Quality, Bing provides users with control over how their data is collected, used, and personalized. Through the Microsoft Privacy Dashboard, users can view, delete, or export their personal data, including search history. They can also adjust advertising and content personalization settings. For example, if a user frequently searches for travel content, Bing might personalize their Microsoft News feed with travel-related stories. These personalization features are optional and can be disabled. In the EU, Bing also requires users to opt in before using cookies or similar technologies that collect data for personalization.

Policies and Other Mechanisms:

Prioritizing user fundamental rights to search: As a search engine, Bing is committed to supporting and prioritizing users' fundamental rights and their rights to access and seek information. Bing does not broadly remove content from its index. Instead, it focuses on minimizing unintended exposure through ranking and targeted filtering. Where content is removed from search, Bing is transparent by notifying users through a notice on the search engine results page and publishing transparency reports.

Classifiers: Bing deploys specialized mitigations in its GenAI features, such as "classifiers," to reduce the risk that harmful content might be generated by Bing. These classifiers analyze text, images, and videos to identify potentially harmful content in search queries and generated responses. By applying AI-based classifiers and content filters, Bing can flag and prevent the production of harmful content, including in

features such as Copilot Search and Bing Image and Video Creator. For instance, classifiers have been implemented to mitigate the risks of generated content related to public figures and to prevent the generation of low-authority content.

Terms & Conditions – Webmasters: Bing is transparent about its policies and actions with respect to webmaster content and makes these documents available in local languages. Bing's Webmaster Guidelines are accessible via the link.

Responsible Product/Service Development:

GenAI – Grounding in High-Ranking Search Results: Copilot Search responses are generally grounded in high-ranking search results, a process known as "grounding." This approach involves providing data to a large language model (LLM) to improve its ability to generate accurate, relevant, and updated outputs. By centering responses on high-authority content from the web, Bing aims to prevent the generation of content containing private information hosted on low-authority sites. Additionally, providing links to source websites allows users to evaluate the credibility of the information presented.

Continuous Improvement of Ranking Algorithms: Bing's algorithms are the first and strongest defense in delivering safe, authoritative content. These systems are designed to match user queries with the most relevant third-party web pages in Bing's index, elevating trusted sources and reducing visibility of unreliable or manipulated content. Rather than removing content outright, Bing relies on continuous improvement of ranking algorithms to ensure that harmful content is not surfaced prominently, particularly when users have not explicitly searched for it. To support this, Bing invests continuously in refining how its crawlers and algorithms detect quality content. These improvements are applied across all markets and languages where Bing operates. Teams regularly assess performance through key metrics, including internal safety indicators, and use these results to inform updates. In addition to technical signals, Bing incorporates feedback from users, Microsoft's policy teams, regulators, and civil society organizations to ensure its ranking decisions reflect evolving standards of trust and accuracy.

Pre-Launch Risk Assessment: Specifically, before any new feature or product is introduced to users, Bing requires it to pass through a robust pre-launch risk assessment. This assessment process is part of Microsoft's broader compliance framework and includes reviews of privacy, security, accessibility, safety, and responsible AI concerns.

These reviews are conducted by professional compliance managers who ensure that products meet Microsoft's standards and legal obligations. Where risks are identified, teams must implement appropriate mitigations before the product or feature can go live. This process ensures that safety, privacy, and integrity are designed into Bing experiences from the start.

Data Protection Impact Assessment ("DPIA"): For any new products or features that pose materially different personal data processing challenges, Bing is committed to completing or updating full DPIA as part of its rigorous privacy protocol. These assessments are thoroughly reviewed by Microsoft's Data Protection Officer and are subject to reevaluation annually or sooner, should there be a significant change to processing risks. This process ensures that Bing's products and services comprehensively identify and address privacy risks at each level of development and deployment.

Cultural Sensitivity Safeguards – Maps: Bing Maps follows requirements for cultural sensitivity or laws to display differing names or geopolitical boundaries depending on the location of the individual conducting the search.

Risk Management:

Safety Measures – Prioritizing High-Authority Content in Search: Bing's commitment to Trustworthy Search Principles means prioritizing high-authority content in search results, especially in areas where inaccurate or manipulated content may pose risks. Rather than removing content outright, Bing relies on ranking algorithms to ensure that harmful content is not surfaced prominently, particularly when users have not explicitly searched for it.

To support this, Bing uses a Quality and Credibility ("QC") score, which includes external signals that help identify authoritative sources. These signals help elevate trusted results while demoting pages that may be misleading or exploitative. Bing also uses ranking systems to defend against manipulation, including spam or search engine optimization abuse that could otherwise distort results.

These ranking systems are continuously measured and refined to ensure they reflect user expectations, policy goals, and emerging risks. The result is a more accurate and safer search experience that remains grounded in user intent.

Product Safety

Compliance Management

Internal Training: Microsoft requires employees to undergo regular training in key areas of legal and policy compliance, including privacy, security, and standards of business conduct. Relevant product managers receive additional training and regular communications from expert teams in their areas of focus, such as for digital safety, privacy, or RAI. Staff with specific safety responsibilities must complete at least one digital safety specific training per year.

Bing leverages the investments made by Microsoft in content reviewer training and wellness - from top notch medical benefits to a robust Employee Assistance Program ("EAP") service, as well as comprehensive fitness and mental health programs on site and virtually. Content Moderation and review teams have guidelines of what not to review, how to address the content they accidentally view, and the ability to opt-out of work related to mental health and wellness. Furthermore, general Microsoft health and wellbeing policies apply to ensure employees' holistic wellbeing – mental, emotional, physical, and financial - is supported.

Content and Activity Moderation

Safeguarding the rights and protection of minors: Bing has implemented a comprehensive suite of measures aimed explicitly at safeguarding the rights, privacy, safety, and security of minors across its service. These measures include setting age restrictions under the MSA, developing age-appropriate privacy communications, ensuring Bing features are accessible and understandable to teen users, as well as enforcing strict advertising policies through Microsoft Advertising. Additionally, Bing's commitment to privacy protection for minors is evident through practices such as prohibiting targeted advertising to users under 18, minimizing data use, and engaging with external stakeholders on children's safety and development. Through these efforts, Bing demonstrates its dedication to creating a secure and supportive online environment for young users, while continuously seeking to enhance its protective mechanisms in alignment with evolving standards and stakeholder feedback, including Microsoft policy teams, regulators, and civil society, to continue to iterate and improve on trust and safety in the Bing service.

Blocklists: Bing Image and Video Creator uses targeted safeguards to limit the generation of inappropriate or potentially harmful content. A key intervention is the use of intelligent blocklists—filters designed to detect

and restrict specific user inputs that contain sensitive or risky names, terms, or phrases. These inputs are sorted into "hard" or "soft" blocklists, depending on the level of sensitivity and risk associated with them.

The development and refinement of these blocklists is a collaborative process involving internal product and safety teams, insights from jailbreak researchers, red teams, and specialized legal counsel. For example, during election periods, Bing consults with elections counsel to ensure the blocklists reflect heightened sensitivities around electoral misinformation or manipulation. This multi-disciplinary approach allows Bing to tailor safeguards to real-world risks in a dynamic and proactive manner.

Age Restrictions for GenAI Features: Bing has certain age restrictions for GenAI features. While some, limited functionality in Bing Image and Video Creator is available to all unauthenticated users, full functionality is restricted to authenticated users who meet the appropriate age thresholds. Users must be signed in with a Microsoft account that reflects an age of 13 or older, or a higher age in jurisdictions where parental consent laws apply, to access full functionality, including multiple turns on the same content. If a user's account indicates they are below the required age, they are prevented from accessing Bing Image and Video Creator, even the limited experiences available to unauthenticated users. This age-based access control helps protect younger users from exposure to content or interactions that may not be appropriate for development.

Prevention of Targeted Advertising to Minors: Microsoft does not deliver personalized advertising to minors whose birthdate in their Microsoft account identifies them as under 18, in accordance with the Microsoft's Privacy Statement Children and advertising. Across Bing features, authenticated users under 18 are not subject to targeted advertising. This important protection for youth will be carried forward into the new Bing features as well. Per the Microsoft Advertising Remarketing policy, advertisers must also not (i) "create a remarketing list, retarget or otherwise profile" any individuals under the age of 18; (ii) target individuals under the minimum age required for the product advertised; and (iii) implement remarketing on sites or applications directed to minors under the age of 18 or other applicable age limitations in an applicable market. Similarly, Microsoft has clear policies in place to prevent showing ads that are behaviorally targeted or contain adult materials such as alcohol or gambling to known child users. Microsoft uses both manual and systematic methods to review ads for compliance with its policies and reserves the right to remove those that are non-compliant.

Privacy Training: Bing has implemented a multifaceted approach to product enforcement to protect user privacy. Central to this strategy is the requirement for Microsoft personnel to undergo annual privacy training. This ensures that each team member is well-versed in appropriate data use practices, reinforcing the company's commitment to safeguarding user privacy.

Data Governance

Safeguarding User Data: Bing was built with privacy in mind, so that personal data is collected and used as needed and is retained no longer than is necessary. Microsoft applies industry-leading data storage policies and practices to ensure the safety and security of user data that is collected and stored. Some of these practices include, but are not limited to, storing data in Microsoft-owned and operated data centers with standard physical security access and the data is encrypted from data subject to data center in transit. The Visual Search feature deploys a mechanism that blurs faces in the images at the time of the upload by users, so that facial images are not further processed or stored. Access to most Bing Search history data is limited to Bing employees. More information about the personal data that Bing collects, how it is used, and how it is stored and deleted is available in the Microsoft Privacy Statement.

Safety Measures – Privacy: Bing's terms of use governing activity on its GenAI features prohibit the use of

the service to violate others' privacy. Safety systems in GenAI features are designed to prevent abuse, such as blurring faces in images before they are used as prompts in visual search and preventing the generation of images with individual faces. Bing labels images as generated by Bing Image Creator to limit the possibility of misuse.

Data Retention and Deletion Policies: Bing Data retention is managed within a set schedule. Bing limits the retention of data to that necessary to provide the service and ensure the safety of the service.

Unauthenticated users can control their search history through Bing's in-product controls using Bing's settings page.

External Engagement

Verified Project: Microsoft has extended its support to the Verified Project in response to disinformation campaigns targeting Slovakia. Launched in the fall of 2023, Verified is a browser extension and digital literacy program that uses AI to transparently rate the credibility of online news articles. With the capacity to process large amounts of data, Verified is able to recognize manipulative elements in the text.

Governance

Microsoft Threat Assessment Center: Bing is actively engaged in enhancing public security through the collaboration of its internal teams, such as the MTAC, and external experts specializing in public safety and misinformation. MTAC is a group of experts who analyze and report on nation state threats, including cyberattacks and influence operations. The team is responsible for detecting, assessing, and disrupting digital threats to Microsoft, its customers, and democracies worldwide. MTAC brings together a team of experts fluent in more than a dozen languages able to analyze both cyber and influence threats arising from nation states and the most prolific threat actors. Detecting emerging actors and methods used in malign influence operations is highly resistant to automated processes, but Microsoft has developed technology to pair the right data with the right team of human analysts. With data and geopolitical expertise on both the target populations and the entities engaging in malign influence, the team can detect new actors and methods—assessing and in many cases attributing influence activity. This cooperative effort is pivotal in swiftly identifying and addressing potential threats, especially in preparation for critical global elections including those in the EU. Through these measures, Bing demonstrates its commitment to maintaining a secure service by mitigating risks that could compromise public security.

Policies & other Mechanisms to Counter Harm

Special Authoritative Elections Answer in Search: Bing launches special informational and Answer segments for elections in defined markets to show high authority information related to the election for users, localized by language and member state. Bing Search directs users to official sources for elections, such as https://elections.europa.eu/, and other high authority sources. Bing also offers an authoritative "How to Vote" answer to government election resources for elections in defined markets.

Safety Features – Visual Search: Bing visual search features (which allow use of images as a search prompt) do not allow for matching of images of private individuals' faces with content in the search index. It also does not allow for searching of adult images. In the visual search feature, Bing blurs faces before processing data contained in files provided by the user.

User Privacy Choices – GenAI: Bing empowers users to make informed decisions about their privacy choices and how their data is collected and used in GenAI experiences. Through the Microsoft Privacy

Dashboard, authenticated users can view, export, and delete their stored conversation history, including any queries or prompts submitted in Copilot Search.

These controls complement existing privacy settings, such as cookie banners and consent mechanisms, which continue to apply to generative features. Bing also applies an internal review process to examine how choices are presented, ensuring users clearly understand its options and the implications of its decisions. By giving users meaningful control over its data, Bing supports autonomy and accountability in the use of AI-driven technologies.

Digital Literacy: To this end, Microsoft promotes digital literacy across its services, including Bing. This work is essential in helping users critically assess online content, especially in the face of misinformation or potentially deceptive material, and complements Bing's other proactive safety efforts.

In Product Indicators – Child Safety: For users search queries for Child Sexual Abuse Material Bing displays PSAs, similar to those regarding negative searches regarding suicide, stating that Child sexual abuse material is illegal, but also includes links to report these illegal materials or receive anonymous support, aimed at those who are victims of CSAM. For image and video queries relating to CSAM, no results are shown, and instead shows unrelated images from popular web content, and for Copilot Search, there is no summarized response to CSAM related queries

Public Service Announcements: For certain high-risk query categories, Bing displays *PSAs* at the top of search results. PSAs highlight authoritative information including reporting and support options, and/or offer clear warnings about potential risks. There are, for example, PSAs in the EU related to the illegal nature of child sexual abuse material and resources and support for queries related to suicide & self-harm. PSAs are one of Bing's most visible tools for helping users navigate sensitive or dangerous topics with clarity and care.

These messages are localized, translated, and adapted to serve users in specific markets where Bing operates. Whether it's a PSA or a warning label, the goal is the same: to provide clear, immediate context when users encounter results that may carry risk.

Microsoft Advertising Policies: Microsoft Advertising, which powers ads on Bing, has clear and regularly enforced content policies and practices that prevent advertisements negatively impacting human dignity. The Microsoft Advertising Policies set out the requirements for ad content, including criteria upon which ad content will be removed. Microsoft requires its advertisers and partners to comply with its policies throughout their use of the Microsoft services. Microsoft Advertising also has a set of Relevance and Quality Policies to manage the relevance and quality of the advertisements that it serves through its advertising network.

Microsoft Advertising has terms of use and content policies are published online and easily accessible for advertisers and Bing users. These terms are updated as needed and provided to users in relevant languages.

Advertisers also retain ownership and responsibility for their ad content, but the advertiser must agree to its terms when signing up for a Microsoft Advertising account.

Microsoft Advertising employs dedicated operational support and engineering resources to enforce these policies, combining automated and manual enforcement methods to prevent or take down advertisements that violate its policies.

Microsoft Advertising restriction on political ads: Microsoft does not allow political advertising within the MS Advertising ecosystem, which supports advertising on Bing. Microsoft Advertising policies prohibit ads for election-related content, political candidates, parties, ballot measures and political fundraising globally; similarly, ads aimed at fundraising for political candidates, parties, political action committees and ballot

measures also are barred. Microsoft Advertising's policies also prohibit certain types of advertisements that might be considered issue based. Microsoft Advertising is a signatory to the COPD and more details on how its commitments is contained in Microsoft's most recent COPD report Bing has partnered with Microsoft Research and third-party research organizations to contribute to novel research and internal studies concerning safe design practices, RAI, and information manipulation. In preparation for global elections, Microsoft Research recently conducted internal research concerning information integrity and elections in the age of GenAI.

Responsible Product/Service Design

Microsoft's RAI Standard: All Microsoft products that incorporate artificial intelligence, including Copilot Search and Bing Image and Video Creator, are governed by Microsoft's RAI Standard. This standard requires comprehensive reviews and mitigation planning before AI-powered features are launched.

These reviews focus on identifying potential risks, assessing the likelihood and severity of harm, and ensuring mitigations are robust, context-appropriate, and in line with Microsoft's AI Principles. For GenAI experiences, this often includes added layers of technical safeguards, transparency mechanisms, and human oversight. By rigorously evaluating generative features before deployment, Bing prioritizes user safety and dignity, while still enabling innovation. This process reflects Bing's commitment to maintaining a respectful, secure digital environment in the face of emerging AI capabilities.

Conversational Drift: To address issues related to conversational drift, Bing has limited the number of turns per chat session in Copilot Search. This limitation helps prevent responses that are repetitive, unhelpful, or inconsistent with the intended tone. The team continues to evaluate additional approaches to mitigate this issue.

Prompt Enrichment: In cases where a user's prompt is ambiguous, Copilot Search and Bing Image and Video Creator may use the LLM to build out more details in the prompt to help ensure users receive the response they are seeking. This prompt enrichment does not rely on any knowledge of the user or their prior searches but instead on the AI model. These revised queries are visible in the user's chat history and can be deleted using in-product controls.

Targeted limitations on GenAI text and image responses: To address risks associated with Bing generating harmful content, Bing has placed specific, targeted limitations on GenAI text and image responses for certain queries. For example, when Bing classifiers identify potentially harmful content in user prompts for Copilot Search and Bing Image and Video Creator, Bing may not return generated content to users, may divert users to different topics, or redirect users to traditional search results. These measures help prevent the dissemination of harmful or inappropriate content.

User-Centered Design: User-centered design and user experiences are essential aspects of Microsoft's approach to responsible AI. The goal is to root product design in the needs and expectations of users. As users interact with Copilot Search and Bing Image and Video Creator for the first time, these services offer various touchpoints to help them understand the capabilities of the system, disclose that they are powered by AI, and communicate limitations. Elements of the experience also help users better understand its interactions, including chat suggestions specific to RAI, explanations of limitations, ways to learn more about how the system works, and easily navigable references to show users the results and pages in which responses are grounded.

In-Product Indicators and Disclosures: Bing's safety philosophy includes keeping users informed when safety-related actions are taken. This includes in-product indicators and disclosures such as warnings,

disclosure footers, or icons that explain when content has been removed or flagged, and why. For example, Bing may include a notification in search results when content has been removed due to legal or policy requirements. In AI-powered experiences, Bing includes clear disclaimers that the user is interacting with an AI system, along with reminders to verify content using authoritative sources. These disclosures are written in plain language and designed to be understandable even by younger or less technical users.

AI Transparency: Transparency is essential in building trust in AI. For AI Transparency, Bing displays prominent disclosures and in-product notices when users are engaging with AI-generated content. For example, in Copilot Search, an information icon near the feedback buttons alerts users that the response was generated by AI. Hovering over this icon reveals a plain-language explanation, with links to more detailed documentation. These disclosures are written at a fifth-grade reading level to ensure they are accessible to young users and audiences of varying digital fluency. To support informed engagement and reduce manipulation, Bing has also introduced content integrity as a Service, a system that uses digital credentials and cryptography to authenticate the origin of media. This helps users trust that the images or media they are seeing have not been tampered with, and that they come from verified sources.

Ongoing Monitoring of Risks: Prior to a product launch and during the software development lifecycle ("SDLC"), feature teams, with the support of Bing trust and safety and compliance teams, as well as cross-company subject matter expert teams, conduct a series of risk assessments as applicable to their feature, including for example, the RAI Assessment Process, One Compliance ("1CS") assessments, Security Assessment, and Privacy reviews. In addition to risk assessments, when any features are being created, metrics are evaluated and updated regularly.

Authoritative Content: Bing actively directs users to trusted sources through several tools. For high-risk events such as various EU elections, Bing highlights answers and PSAs that point to expert-vetted information. During events like Public Health crises, Bing would apply defensive search interventions such as demoting low authority results in ranking on queries that may lead users to data voids and misleading content informed by threat intelligence.

Safety Measures – News: Bing prioritizes freedom of information and pluralism for search, allowing a greater variety of results and news entities to appear when users search for them, however for the News feature Bing is more restrictive and seeks to recommend content that meets Bing News Publisher Guidelines. Bing balances this prioritization by working with in-market teams to identify high trust news sources for content that is returned in results for users queries and attaching Content Credentials to help users identify the reliability of the sources they are accessing.

Periodic Defining of New Metric Targets: Bing evaluates the need for resource allocation to address emerging concerns. Bing sets formal OKRs every six months based on review of metrics and areas for improvement, and they consider engagement across the organization as well as investment levels as part of the OKRs. On a monthly basis, Bing senior leadership aligns with cross-functional teams to ensure resources are allocated appropriately for especially key areas of content concern. Finally, Bing sends externally facing daily updates on key areas of content and conduct concern to ensure appropriate prioritization particularly for improvements to GenAI features.

Red Team Testing: Bing also requires extensive red team testing to explore safety risks under pressure. These tests simulate how systems might be misused or manipulated in the real world. Multidisciplinary teams and at times, invited external experts, rigorously test how Bing Search, Copilot Search, and Bing Image and Video Creator respond to adversarial inputs or edge cases.

Red team findings are used to refine training data, harden safeguards, and improve system behavior. Bing

ensures that these tests continue even after launch, maintaining vigilance as systems evolve.

Shiproom review: Every change to Bing, whether in core search or newer features like Copilot Search, must also pass a "shiproom" review. This structured launch-readiness evaluation ensures proposed updates meet rigorous standards across safety, privacy, security, and accessibility.

Shiproom reviews require product teams to demonstrate the impact of proposed changes on Bing's core safety metrics, such as DDR. Teams must provide evidence of both offline and live (A/B) testing and explain how they've addressed any regressions or new risks uncovered during testing. Further, in the monthly shiprooms, issues and risks are presented to Bing leadership, enabling leadership ongoing governance and oversight. These checkpoints ensure that safety remains a central consideration in how Bing evolves.

Risk Management

Feature Evaluation: Bing relies on its extensive engineering infrastructure, defensive search, and quality assurance teams to ensure that new features and metrics are evaluated routinely and are functioning to design.

Targeted Interventions

Content and Activity Moderation

Safety Measures – Defensive Search Interventions: As part of the Safety Measures – Defensive Search Interventions mitigation, Bing has a dedicated team accountable for implementing algorithmic defensive search interventions (as well as metrics monitoring and remediation) to address high impact issues in search results, such as information manipulation, hate speech, and other problematic content that could negatively impact user safety.

Bing uses query classifiers to identify queries that may be associated with elevated risk, such as those that could surface degrading, harmful, or exploitative content, and those trigger a set of defensive search interventions. These algorithmic measures are integrated rather than standalone and work together to uphold Bing's relevance, quality, and credibility standards—particularly in cases where harmful content may exploit data voids or manipulate user intent.

When a query is flagged as potentially risky, Bing may respond with measures such as boosting high-authority sources or applying algorithmic dampening to deprioritize low-quality or harmful content. These interventions are applied in accordance with Bing's core search principles and are actively monitored by the dedicated team that manages implementation and assesses impact across high-risk areas. This combined approach allows Bing to detect potentially harmful search patterns early and respond with targeted mitigations that preserve both user safety and access to high-quality information.

Safeguards – Shopping: Shopping uses a two-filtering system where Ads team conduct a first vetting and Bing conducts a second vetting process (Content 95% comes from Ads and 5% comes from Shopping crawl data).

Social Listening: Bing maintains social listening pipelines where insights and user feedback on Bing's features are collected from the Internet. These insights and user feedback are manually reviewed, analyzed daily, and shared across Bing's product teams and product engineering leadership in a daily report. These reports inform Bing as to the public concerns, issues, or emerging trends and can serve as barometer of public sentiment on various topics related to Bing.

Data Governance

Copyright Infringement: Bing encourages respect for Intellectual Property ("IP") rights, including copyrights, while also recognizing the rights of users to engage in uses that may be permissible under applicable copyright laws. If Bing receives a legally sufficient notice of copyright infringement from the copyright owner or its authorized agent, Bing may remove from its search results links to webpages containing material infringing the rights of the owner of copyrighted content. If a rights holder has an IP concern about a website's content that is linked to by Bing, or a Bing ad, they can review and contact Bing through the Report Infringement page. In certain circumstances relating to quality, safety, and user value, Bing may decide to remove certain results or may warn or educate users or provide options for tailoring results.

External Engagement

Redirecting Search Queries – Copilot Search: For Copilot Search, queries related to self-harm or suicide are considered high-risk, and Copilot Search responds to these queries through "thoughtful disengagement" and redirects users to search to reach out to family members, friends, and mental health professionals. For search queries regarding eating disorders, under Bing's content policies this falls under a form of self-harm, as it can be considered a physical risk for users, thus defensive interventions for user queries are employed to direct users towards high authority content. For responses generated by Copilot Search, there is a high authority approach taken regarding queries relating to eating disorders, to explain the negative effects of certain associated behaviors.

External Partnerships related to Public Security: Bing takes a multi-stakeholder approach by having External Partnerships related to Public Security. Additionally, as a member of the Global Internet Forum on Counterterrorism (GIFCT), an initiative to bring together technology companies, governments, and experts, to share means of countering terrorist and violent extremist from exploiting digital platforms, Bing has access to the GIFCT's Incident Response processes, including ingesting hashes related to an event activated as Content Incidents or Content Incident Protocols. This allows Bing to quickly become aware of, assess, and address potential content circulating online resulting from a terrorist or violent extremist event.

Bing also participates in the Advisory Network (a multistakeholder advisory group) of the Freedom Online Coalition, a coalition of 37 governments working to advance Internet freedom and works closely with the ISD to understand risks related to TVEC in search, as well as to surface "counter narratives" via ad grants that help deter users who indicate an intent to learn more about extremist organizations.

Microsoft also supports reputable third-party credibility ratings of online news, and Bing has been a leader in the use of trustworthiness indicators in search to help users evaluate the credibility of news sources they encounter.

Partnership on AI: Microsoft, including key members of the Bing Search team, actively participates in the Partnership on AI ("PAI"), collaborating with peers across industry and civil society to identify and mitigate potential harms associated with synthetic media. As part of this work, Microsoft has contributed to the drafting of the PAI's Synthetic Media Code of Conduct; which are guidelines that promote the ethical and responsible creation, distribution, and use of AI-generated content, including deepfakes and digitally altered media.

Microsoft Threat Intelligence for Nation State Information Operations: Microsoft works to identify and track nation-state information operations targeting democracies across the world and works with trusted third-party partners, to provide early indicators of potential threats that can be leveraged to inform early detection and defensive search strategies.

Governance

Threat Intelligence: Bing has a Threat Intelligence team that specializes in identifying and responding to high-risk search topics. This team continuously investigates where and how search results might be falling short and applies corrections with precision. The team toolkit includes boosting the visibility of authoritative websites, demoting those with low reliability, and restricting suggestions that could lead users toward harmful or misleading content.

Policies & Other Mechanisms to Counter Harms

Directing to Authoritative Sources – Search: In cases where Bing detects threats to search quality in fast-moving or sensitive areas, Bing will direct users to authoritative sources. This includes areas such as elections, public health, or active manipulation of the information environment. Bing may demote low-authority content, boost more credible sources, restrict problematic autocomplete suggestions, or make manual adjustments to search results to limit harm. Each step is narrowly applied, with the goal of improving user safety while preserving open access to information.

User education – GenAI search results: Bing offers tools to help users understand the context and trustworthiness of search results. Bing works to be transparent about how its safety systems work for GenAI features, and for novel systems such as Copilot Search reminds users that they are engaging with an AI system, and to remind users of the importance of verifying facts in the source material. Bing ensures that users can find third-party links to content responsive to their queries in standard search results.

Webmaster Abuse Monitoring: Bing's Webmaster team invests significant resources in addressing attempts by webmasters to engage in abusive Search Engine Optimization ("SEO") techniques and Bing's ranking systems are designed specifically to prioritize high authority content, rather than websites deploying spam or other manipulative tactics. Bing teams regularly monitor trends and emerging threats to evolve mitigation practices to meet the rapidly adapting fraud, mainly manifested by spammers attempting to manipulate search results and Copilot Search responses. Additionally, Bing's general abuse/spam policies, detailed in Bing's Webmaster Guidelines, prohibit certain practices intended to manipulate or deceive the Bing Search algorithms.

The primary content that appears on Bing is information from third-party websites that is linked in search results. Webmasters are not users of Bing and Bing has no contractual privity with webmasters, so Bing has limited ability to enforce rules regarding content appearing on those third-party sites. However, Bing has still taken steps to be transparent to users and webmasters as to how it ranks and moderates web content via published documents.

Responsible Product/Service Design

Limit Lengths for Generative Search Queries: Copilot Search is limited by the 2000 character query length limit.

Bing's Safety Playbook: Bing has a defined set of review protocols, including policies for handling reported violations and escalations in Bing's Safety Playbook. Bing's Content Moderation team adheres to the guidelines in the playbook in handling violations and escalations.

Legal Review: Microsoft legal teams evaluate product features and proposals to ensure compliance with applicable laws and address potential safety and legal risks posed by a product or feature. Microsoft legal teams include experts in privacy, human rights, AI, technology, marketing, IP, digital safety, consumer protection, information integrity, and litigation, among other areas and provide holistic support to Microsoft

product teams.

Risk Management

Safety Measures – GenAI – Elections: Novel elections classifier intended to restrict Copilot Search from responding to certain election-related queries out of an abundance of caution. This classifier has been a focus of continued improvement, including focused testing in key EU languages used by Bing users in EU.

AI Generated Content Detection: Microsoft is harnessing the data science and technical capabilities of its philanthropic research center focused on social benefits of AI as well as Microsoft's threat intelligence for nation state information operations to better detect abusive synthetic media and other harmful content on the Internet. Furthermore, Microsoft's digital crimes function is investing in new threat intelligence work to pursue the early detection of AI-powered criminal activity.

Continuous Monitoring of Metrics: Bing regularly reviews the efficacy of its interventions to ensure that they are performing as expected and not inadvertently introducing additional bias or other harm. Bing monitors metrics, incidents, and potential risks in the product via red teaming, social listening, and other intake channels. Bing's mitigations are designed to work across languages and markets where Bing is offered. While the robustness of the mitigations might vary depending on the traffic volume, Bing works to ensure the effectiveness of the mitigations by monitoring the metrics.

Content Moderation – Webmaster: Bing monitors for violations of its Webmaster Guidelines (e.g., improper attempts to manipulate search algorithms via keyword stuffing or other prohibited practices) and terms of use, and actions violations consistent with its policies and procedures.

Monitoring Emerging Threats: Bing regularly monitors trends and emerging threats in order to evolve mitigation practices to meet the rapidly adapting tactics to circumvent Bing safeguards and access potentially harmful content.

Content Moderation – Advertising – Continuous Improvement: Microsoft Advertising continually seeks to add new signals and information sources to improve its detection systems, evaluating the incremental performance of adding these signals, and sources to sweeps and models and adding those that it finds useful to the system.

Content Moderation

Compliance Management

Content Moderation – Training and Wellness – Child Safety: Bing's Content Moderation Ops team carries out training for content reviewers to ensure alignment on policy, labelling, and enforcement. For high-consequence harms, like child sexual exploitation and abuse, specialized teams receive additional focused training and wellness resources.

Content and Activity Moderation

Content Moderation – Search Content Removal: Bing maintains rigorous and principled processes, policies, and procedures for content removal and maintains policies and other mechanisms to counter harms. Central to these processes is a firm commitment to the prevention of CSEAI. Bing proactively scans content to detect known CSEAI using hash-matching technologies such as PhotoDNA and MD5. This scanning occurs before content enters the search index and is used to ensure such material does not appear in results at all, across Core Search and Copilot Search. Bing also proactively scans upload image uploads in

Bing Image Creator and Bing Video Creator to prevent known CSEAI from being uploaded. This proactive detection is supplemented by intelligence from trusted partners, including NGOs and law enforcement, and through user reporting channels. When CSEAI is identified, Bing removes it globally and without hesitation. This is a zero-tolerance area where Microsoft's commitment is absolute.

Bing also works to remove adult or sexually explicit images of another person online when shared without that person's consent. This is commonly referred to as nonconsensual intimate imagery (NCII), or "revenge porn". Microsoft considers this to be a gross invasion of personal privacy – whether the content is real or a "deepfake" image that is created using AI or other photo editing tools

Finally, Bing also removes certain other content under global legal standards or Microsoft policies, such as materials containing extremely sensitive personal information that could be exploited for fraud or identity theft. Removals may be prompted by legal orders from governments, user-generated reports, or policy-based decisions following internal investigations. Microsoft applies rigorous standards to each request, including the potential impact on user rights and public interest.

Report a Concern: Bing provides users with an accessible report a concern mechanism on every page of Bing Search and across GenAI experiences. Through this function, users can raise concerns about potentially unlawful or harmful content. These reports are routed to human moderation teams who assess submissions based on Microsoft's policies, relevant laws, and the specific context of the reported material. While not all reports lead to content removal—especially in cases involving lawful but offensive or disputed content—each report contributes to Bing's broader understanding of emerging risks and informs future updates to its safety interventions.

Account Suspension – Advertising: Bing reserves the right to issue immediate suspensions of advertiser's accounts should they egregiously violate Microsoft's terms or policies, if those violations intend to use Microsoft services in a fraudulent, harmful, or illegal manner. Microsoft uses both manual and systematic methods to review ads for compliance with its policies and reserves the right to remove those that are non-compliant.

Content Moderation – News – Proactive Detection: Bing scans News content using standard Bing classifiers to flag content for action, and to filter out lower quality sources. There is a continuous process of optimizing classifiers for precision, recall, coverage, and reducing defects without any significant fundamental ranking architecture changes. Bing emphasizes presenting varied perspectives in search results, particularly valuing media pluralism within Bing News by enforcing strict publisher guidelines and promoting reliable news sources. Bing's nuanced approach ensures that, except in cases where content is illegal or violates another fundamental right such as privacy or human dignity, information remains accessible. For government demands, Bing employs additional safeguards to ensure any actions taken are narrow, specific, submitted in writing and based on valid legal orders.

Social Listening for GenAI Features: Bing maintains a social listening pipeline where insights and user feedback on Bing's GenAI features are collected from the open Internet, from users and potential users. These insights and user feedback are manually reviewed by humans, analyzed daily, and shared across Bing's product teams and product engineering leadership in a daily report, to ensure the insights feed into product evaluation and further improvements. These reports inform Bing as to the public perception of mitigations and serve as a barometer on topics concerning Bing's operations.

- Microsoft proactively tracks tactics and methods that users employ to jailbreak Copilot Search and Bing Image and Video Creator to produce content with a negative impact on user safety and takes daily action to address any gaps in safeguards.

- Microsoft regularly reviews the prevalence of issues within Bing Image and Video Creator to evaluate the landscape of potential harms to human dignity and adjust the safeguards accordingly. Through this process, Microsoft enhances the mitigations to reduce the occurrence of harmful outputs on human dignity.

Content Moderation – Tooling: Bing utilizes its own Content Moderation Platform ("CMP"), which is an internal tool for implementing content moderation decisions, such as URL exclusion and query treatment. Bing's Content Moderation team uses CMP to effectively take actions on URLs/queries based on the Bing Search policy and implement the decision in the backend.

External Engagement

Content Moderation – Content Removal Policies and Procedures: Bing will remove materials identified as illegal based on practice scanning, in limited cases, and following investigations upon notice.

Bing proactively scans content for CSEAI using PhotoDNA. Bing proactively scans content for CSEAI using hash-matching technologies (including PhotoDNA and MD5) to detect matches to known CSEAI, to avoid it from appearing in the search index. Bing also relies on reactive reporting as well as threat intelligence provided by third-party expert partners to identify and remove CSEAI from the services.

Bing also takes steps to globally remove (based on Bing policies) certain other categories of content that may be illegal materials in some jurisdictions, such as extremely sensitive personal information that could be used for identity theft or other fraud. Bing may remove content based on legal removal orders, government takedown requests, copyright infringement removal requests, Right to be Forgotten requests, and user reports concerning the content appearing in Bing Search, GenAI Experiences, and Ancillary Search Features.

If Bing receives requests to remove content from individuals, businesses, and governments, in limited cases, where quality, safety, user demand, relevant laws and/or public policy concerns exist, Bing might remove results, inform users of certain risks through PSAs or warnings, or provide users with options for tailoring their content.

Right to be Forgotten: Bing's commitment to privacy and dignity is also reflected in its implementation of the Right to Be Forgotten ("RTBF"). Under this policy, users in jurisdictions such as the EU can request the removal of search results associated with their name if those results are outdated, inaccurate, irrelevant, or excessive. Requests are reviewed by trained staff using clear criteria to balance the individual's right to privacy with the public's right to access information. Where appropriate, Bing removes URLs and associated search suggestions from the index in response to validated requests. Escalation pathways are in place for complex cases, including collaboration with local courts and data protection authorities where necessary.

This process is underpinned by strong training protocols and quality assurance systems. Bing ensures consistency and fairness in RTBF decisions, while maintaining transparency about the process through user guidance and regular engagement with regulatory authorities.

Content Moderation – Travel Stories: Travel Stories are provided and moderated through a third-party partnership. This party maintains content moderation and quality assurance processes that uphold Bing's standards. Licensors commit that the data content is not misleading, false, libelous, defamatory, obscene, unlawful, or injurious to any third-party. Bing undertakes additional quality checks of Travel Stories content, and Bing can report issues to the vendor to remediate. Some of these data content issues include inaccurate URLs, dead links, adult, risqué or deleterious data content, copyrighted data content, data content that violates Microsoft policy, or illegible characters. For first priority items, third-party response time is less than 24 hours and resolution time is less than 3 days and corrected feed or Application Programming Interfaces

("API") content is delivered within 5 days after first reported.

Content Removal – Legal Reasons: Certain countries have laws or regulations that apply to search service providers and require search engines to remove links to certain indexed pages from search results. Some of these laws allow specific individuals or entities to demand removal of results (such as for copyright infringement, libel, defamation, personally identifiable information, hate speech, or other personal rights), while others are administered and enforced by local governments. When Microsoft receives a legal request or demand, Bing aims to balance its support for freedom of expression and for free access to relevant content with compliance with local laws. Bing reviews and assesses the request or demand, including the reason and basis for it, the authority or rights of the requesting party, Bing's applicable policies, and Bing's commitments to its users regarding freedom of expression, human rights, and freedom of information. Bing then determines whether and to what extent access to the content should be removed and takes action.

Responsible Product/Service Design

Notice to Users: Bing's approach for Notice to Users is designed to provide notice to users whose content or conduct is at issue in an enforcement action (with relevant exceptions, such as legal prohibition or prevention of further harm).

In many cases, Bing's moderation decisions will not affect user content since Bing does not allow users to upload or share content on the service and there are limited cases in which a user's activities on the service could trigger enforcement actions.

- Webmasters (who are not Bing end users) have the ability to see how their content has been indexed by Bing in their Bing Webmaster Dashboard.
- In the case of Terms of Use violations in GenAI features, users receive in-product notice informing them of access restrictions or suspensions.

Content Moderation – Advertising – Proactive Detection: The Microsoft Advertising detection system is multi-tiered, with checks based on main line ("ML") models, heuristic sweeps/rules, and manual reviews to prevent bad actors from coming onto the system. Overall, Bing has models and rules that apply at the advertiser-level as well as

those that apply to a particular ad.

- Some of the models that apply at the advertiser-level include:
    - Advertiser Identity Verification (AIV): designed to enhance user trust and safety by requiring advertisers to prove their legal identity before they can run ads. This program acts as a deterrent to bad actors who might attempt to damage the network through various malicious activities such as phishing scams, personal information fraud, and advertising disallowed products.
    - Payment fraud detection: This model triggers when the advertiser adds a payment instrument. This model is owned by the commerce risk team and Bing consume its decision.
    - Purchase risk model: This model triggers billing the advertiser to determine whether to try to make the charge or block the advertiser. This helps to bring up-to-date information to determine payment risk.
    - Behavioral risk model: This model looks at various features of an advertiser's account or campaign(s) such as the ad text, keywords, budgets, account age, account structure, attributes of the payment instrument, individuals associated with the account, devices from which the advertiser is accessing the system etc. to determine the risk of an advertiser being fraudulent. This model runs when the advertiser makes any changes to their ad campaigns

(ads/keywords/budgets, etc.).
    - LLM-based spoofing detection: One of the large categories of fraud is where advertisers pretend to be a legitimate site (e.g., trying to deceive users into believing that an ad leads to the official website of a bank). Bing has an LLM-based approach looking at advertiser demand to try and predict if advertiser is engaging in this behavior.
- Bing applies models to individual ads based on the corresponding Microsoft Advertising Policies applicable to the market where the ads serve. Such models include, among others: Adult, Gambling, Tobacco, Weapons, Disallowed Supplements and Health Products, and Sensitive Topics (hate speech, violence, suffering, gore.)
- Based on the above models' output, Bing decides to either close the advertiser or send it to manual review for a decision. In addition to the models described above, Bing evaluates and uses other signals from Microsoft internal and external providers. For example, Microsoft Advertising uses a signal from the Bing team that gives an indication of whether a particular domain has spam. Bing uses signals from the Defender team which indicates that the user would be harmed by visiting this page.
    - Some of the signals from external providers include:
        - Signals from VirusTotal, an aggregator of malware risk signals from multiple providers.
        - Domain tools – which have indicators of domain risk.
        - Who is – which provides useful information such as age/expiry of domain, clicks on the domain on Bing and Google.

Content Moderation – Shopping: Products in Shopping, primarily leverage product advertisements from Microsoft Advertising. Microsoft reviews each listing added to the Microsoft Merchant Center product feed. Product listings that do not comply with Microsoft Advertising Policies are disapproved. Any merchant that violates Microsoft Advertising Policies will also be disapproved, in this case no products from the merchant will be visible. Bing's Shopping team conducts another layer of review on top of Microsoft Advertising when including products within Shopping search results.

Risk Management

Content Moderation – Advertising: Understanding the critical importance of maintaining a trustworthy ad ecosystem, Microsoft Advertising enforces stringent content policies that strictly prohibit misleading, deceptive, and fraudulent advertisement content. This is complemented by a set of Relevance and Quality Policies aimed at ensuring the served advertisements are not only legal but also relevant and of high quality, thus preventing advertisers from utilizing questionable or misleading tactics to lure users. Advertising employs a combination of human and automated review processes using algorithmic systems and a robust filtration system to detect and mitigate harmful online behavior such as fraud, phishing, malware, or compromised accounts. Bing users can also report advertising for takedown through the "feedback" function. Detection of violative ad content covers ad text, images, extensions product feeds and components, landing pages, URLs, keywords, and targeting settings. Malware specifically is considered an egregious violation and will result in the suspension of an advertising account.

Report a Concern – Advertisements: Users are encouraged to report advertisements that promote illegal content or violate Microsoft Advertising terms through a structured feedback form available on Bing's pages. Specific forms for reporting low quality ads or IP concerns are also available.

Incident Response

Content and Activity Moderation

Safety Measures – Family Safety Settings: Parents can use the Microsoft Family Safety Center to allow parents to track their family's search history and monitor screen time. Child users whose parents have access to their search history are aware when these controls are in place.

Incident Response and Service Level Agreements (SLA): At the core of Incident Response and Service Level Agreements (SLAs) is a well-established infrastructure for responding to notices involving illegal or policy-violating content. This includes materials such as CSEAI, NCII, exposed personal information, and other high-risk violations. When Bing receives a credible report from governments, users, or trusted third parties, it acts swiftly to evaluate and address the concern.

Bing's support teams are trained in legal and policy standards, with clear procedures for escalating complex cases to local legal experts and subject matter specialists. Reports are tracked through Microsoft's internal incident management system, which ensures accountability and drives collaboration across engineering, legal, safety, and policy teams. Alerts are sent to designated stakeholders, and all incidents are triaged according to priority, with time-sensitive issues handled under strict SLAs that form part of Bing's broader Objectives and Key Results (OKR).

The prioritization process considers the nature of the report, the sensitivity of the policy area, and contextual details such as region, language, and media format. While response times vary based on these factors, most reports involving high-priority concerns are resolved within 24 hours. This fast-paced yet thoughtful process ensures that Bing's response mechanisms remain both effective and user-respecting.

Regular testing and quality checks are in place to ensure that these workflows operate reliably. Bing's commitment to continuous improvement means that incident response systems are refined over time based on lessons learned, stakeholder feedback, and evolving safety challenges.

Suspension of User Accounts: In addition to its core search product, Bing also supports services such as advertising that involve direct user participation. In these contexts, Bing enforces clear standards to prevent misuse that could harm users or compromise the integrity of the service. Advertisers who engage in fraudulent, misleading, or illegal behavior are subject to immediate suspension of advertiser accounts.

Bing applies these actions judiciously and in alignment with Microsoft's broader safety commitments. Suspension decisions are made based on evidence and internal enforcement guidelines, with appropriate documentation and oversight to ensure consistency and due diligence.

Appeals: Because Bing is a search engine, not a content-hosting service, its users do not post content directly to the service. As a result, Bing does not maintain user-facing appeals processes for content removals. However, where Bing applies enforcement actions, such as when removing or downranking search results in accordance with policy or legal obligations, it offers webmasters a means of appeal.

Through Bing Webmaster Tools, site owners who believe their content has been incorrectly impacted can request a review. These appeals are evaluated by trained reviewers, supported by escalation paths to experts where needed, to ensure outcomes are consistent with Bing's principles and commitments to fairness, transparency, and trust.

Account Restriction and Appeal Process – Bing Image and Video Creator: In serious cases, Bing may restrict a user's access to generative services or suspend a user's account when there is evidence of repeated or intentional misuse. This includes attempts to bypass content safety protections or efforts to generate highly sensitive material. User's who demonstrate clear intent to create CSEAI will have their accounts suspended.demonstrate clear intent to create CSEAI will have their accounts suspended.

For Bing Image and Video Creator, users whose prompts are repeatedly flagged for policy violations may be automatically suspended for a defined period or permanently banned in severe cases. To support fairness and transparency, these users are given an opportunity to appeal such decisions.

Appeals are reviewed by trained human moderators who consider the user's behavior, applicable terms, and context. Based on the outcome, the enforcement decision may be upheld, or the user's access reinstated. This process ensures that even enforcement actions within GenAI experiences are guided by a commitment to procedural fairness and proportionality.

Research and Partnership

Content and Activity Moderation

Partnerships to Identify Potential Threats: Bing partners with trusted independent third parties, such as Non-Governmental Organizations and threat intelligence providers, to inform defensive search mitigations.

Collaboration with third parties and adherence to external evaluations further reinforce Bing's commitment to its policies. By working with external partners to ensure the quality and reliability of content, Bing strives to maintain a high standard for the information it surfaces. Moreover, participating in initiatives like the GNI, which audits Bing, and undergoing independent assessments by organizations like Ranking Digital Rights, Bing subjects its practices to scrutiny, ensuring they align with global standards for Freedom of Expression and Information.

External Engagement

Microsoft-wide Election Support: An important component of the dedicated cross-functional team focused democracy and elections team's strategy is to establish lines of communication between election authorities in the EU member states and to identify risks to electoral processes, including possible foreign information operations targeting elections.

While these broader engagements led by dedicated cross-functional teams focused on democracy and elections are not specific to Bing (and are part of a whole-of-company approach), they can support and inform risk mitigation on the Bing service. Recent measures include:

- Election Communications Hub: Microsoft created and provided access to a new "Election Communications Hub" to support democratic governments & political parties around the world as they build secure and resilient election processes. This hub has provided election authorities access to rapidly report any issues or concerns to Microsoft security and support teams in the days and weeks leading up to their election, allowing them to reach out and get swift support if they run into major security challenges.
- Campaign Success Team: Microsoft helps political campaigns navigate cybersecurity challenges and the new world of AI by deploying a new "Campaign Success Team." This team advises and supports campaigns as they navigate the world of AI, combat the spread of cyber-influence campaigns, and protect the authenticity of their own content and images.
- Specialized Training: Microsoft invests in political group awareness campaigns and trainings, such as Deceptive AI and Elections Trainings for political parties and campaigns across the EU to learn how to identify the deceptive use of AI and misinformation in elections and report it with Microsoft's Deceptive AI Reporting Tool (discussed below).
- Content Credentials as a Service: Microsoft also launched a broader Content Credentials as a Service

("Content Integrity") to enable political candidates around the world to digitally sign and authenticate media using the Coalition for C2PA digital credentials. Though this service is Microsoft wide, it helps reduce content provenance with political candidates.

- Deceptive AI Election Reporting tool. During the Assessment Period, Microsoft launched a new web page (available here) where political candidates can report a concern about a deepfake of the candidate on Microsoft services. This empowers political candidates around the world to aid with the detection of harmful deepfakes. This web page is available, localized throughout the EU.
- Cyber-security and cyber-influence:
    - Cyber-influence Reporting: Microsoft identifies, tracks, and reports on potential information operations and foreign malign influence operations working with partners such as Spanish news agency EFE and through the MTAC. MTAC's first report, "Protecting Election 2024 from Foreign Malign Influence" was released November 2023, providing a baseline for the following election season, including reflections on previous election influence efforts.
    - Cyberattack Defense: Though threats to democracy exist, the tactics of adversaries are constantly evolving. Microsoft is protecting open and secure democratic processes by
    - providing services and technology to secure critical institutions, protect electoral processes from cyberattacks, and build public trust in voting procedures.
    - Security Guidance: Microsoft works closely with its elections-related customers to provide security guidance and tools to improve its cyber-resilience and protect the integrity of the electoral process.
    - International Foundation for Electoral Systems ("IFES") and NDI Cybersecurity: Microsoft is supporting IFES to strengthen the cybersecurity practices of investigative journalists reporting abuse of state resources in elections. Microsoft is also partnering with the NDI.
    - NDI to strengthen the cybersecurity support infrastructure for political parties and campaigns internationally

Researcher & Academic Support: Bing's approach to Researcher & Academic Support which are designed to create processes for supporting academic and other researchers working on relevant subject matter (to the extent permitted by law and consistent with relevant security and privacy standards, as well as business considerations, such as trade secrets).

Microsoft's commitment to fostering a trustworthy digital ecosystem is evident through its strategic partnerships, innovative research initiatives, and dedication to public well-being. Microsoft collaborates with esteemed research and academic institutions to help ensure the integrity and reliability of the content made available through its services.

- Collaboration with Princeton University: Microsoft has collaborated with Princeton University to create a hub for researchers to access social media data, aiming to enhance the identification and tracking of information operations, with global availability including Europe.
- Research and Public Data Access: Bing offers APIs and datasets for public use and research, with no formal requests for data access under DSA Article 40.
- Other Resources for Researchers: Microsoft provides a variety of datasets and tools for researchers, including the MS MARCO and ORCAS datasets, and promotes RAI practices with tools like the mitigation's library.

Microsoft Research: Bing and Microsoft, including through its research arm Microsoft Research, have collaborated with and third-party research organizations to contribute to novel research and internal studies concerning safe design practices, RAI, and information manipulation. Bing is evaluating additional research

data sharing opportunities pertaining to elections.

External Engagement: Both Bing and specialized cross-Microsoft teams regularly engage with external stakeholders in areas of key policy priorities, such as violative content, information integrity, RAI and AI-specific risks, minors, and technology, to help ensure that Bing internal policies, practices, and standards are addressing key concerns of these stakeholders. These engagements can inform the processes of identifying, assessing, and mitigating risks across Bing and provide greater understanding of concerns related to the Bing service and broader societal and technology related issues. External engagement gives Bing opportunities to hear feedback from a range of civil society, non-profit, researchers, and government stakeholders and learn from others in the industry.

External Engagement – Child Safety: Microsoft routinely engages with external experts on child online safety, through bilateral conversations and in forums such as the Internet Governance Forum (IGF). For instance, Microsoft is a founding and active member of the Tech Coalition, focused on facilitating industry cooperation to address online child sexual exploitation and abuse, thus demonstrating a strong commitment to collaborative efforts in combating evolving CSEA challenges, where Microsoft routinely receive insights and information from other industry participants, as well as briefings from researchers. Additionally, Microsoft participated in the Civil Society Dialogue FY24 and explored two key topics; state privacy legislation, data of minors' and AI, with civil society partners. Microsoft presented some draft guiding principles and foundational capabilities for developing AI products and services for young people.

Finally, feedback from external engagements has provided valuable insights into enhancing the user experience for young audiences. These include suggestions for improving geolocation accuracy and the readability of responses, furthering Bing's commitment to transparency and reliability in its digital offerings.

Regulator Engagement: Bing regularly meets with regulators around the world to understand key concerns, share information, and incorporate feedback into product design and safety systems as appropriate.

Societal Resilience Fund: Additionally, together with Open AI, Microsoft has established a Societal Resilience Fund to promote AI education and literacy, supporting organizations like AARP and International IDEA.

Researcher Program to enable EU Researchers: Bing operates a qualified researcher program to enable EU researchers to easily request access for publicly accessible Bing data from a singular landing page. In addition, Bing is continuing to evaluate additional data sources and tooling to support the research community and looks forward to supporting the DSA vetted research program consistent with the recent Delegated Regulation.

Governance

Dedicated Cross-Functional Teams of Experts: Bing collaborates closely with dedicated cross-functional teams of experts across Microsoft. These partners specialize in areas such as digital safety, election integrity, and RAI. They help develop and maintain company-wide policies that guide Bing's decision-making on safety interventions and mitigation strategies. Their input ensures that Bing is aligned with evolving regulatory expectations and public interest standards.

Dedicated Expert Teams – Hate: Bing employs a specific team that is dedicated to deploying targeted algorithmic interventions to reduce the prevalence of hate speech, harassment against protected groups, and other problematic content in search.

Transparency and Reporting

Data Governance

Data Retention and Deletion Policies – Copilot Search: Copilot Search has data retention and deletion policies to help ensure that personal data collected through these features is kept as long as needed.

Transparency Reports: Where content is removed from search, Bing is transparent by notifying users through a notice on the search engine results page and publishing regular transparency reports (available on Reports Hub | Microsoft CSR). Microsoft's Reports Hub is a publicly available source where users can access various transparency reports in areas where Bing reports on various practices. These key areas include RAI, content removal requests (such as copyright or digital safety), privacy (such as "Right to be Forgotten") and security (such as Government Requests.) The Reports Hub contains additional transparency reports such as jurisdictional, community and privacy and security transparency reports that users can easily access. As a result of the conduct of the Systemic Risk Assessment, each year Bing identifies specific areas for focused enhancements in the coming year.

User Transparency: Bing provides users more detailed information on the main parameters it uses for ranking in the How Bing Ranks Your Content section of the Webmaster Guidelines, specifically calling out that in measuring the "quality" of a website, "pages that call for violence, name-calling, offensive statements, or use derogatory language to make a point are generally considered low quality." Bing works to ensure these resources are available in relevant languages and markets.

External Engagement

EU Code of Conduct on Disinformation ("COCD"), formerly known as the Code of Practice on Disinformation: Bing is a signatory to the EU COCD, formerly known as the Code of Practice on Disinformation, and subscribes to many commitments in the code dedicated to reducing disinformation risks, including providing users with indicators of content provenance, supporting good faith research into disinformation by enabling researcher access to data, and providing biannual reporting pursuant to the code on systems and policies in place to prevent disinformation on Bing across the EU, among others. Bing also provides enhanced reporting on agreed upon elections and other declared "crisis" under COPD.

The Tech Accord to Combat Deceptive Use of AI in 2024: Microsoft has worked with a number of other leading technology companies to create the Tech Accord to Combat Deceptive AI in the 2024 Elections. The Tech Accord's commitments are designed to make it more difficult for bad actors to use legitimate tools to create politically related deepfakes and other deceptive AI elections content, while simultaneously simplifying the process for users to identify authentic content. The Tech Accord calls on companies that generate AI content and those that distribute it to strengthen the safety architecture of its AI services by assessing risks and enhancing controls to help prevent abuse. Microsoft is one of the founding members of the Tech Accord, and Bing, responsible AI, and GenAI teams have worked to implement Tech Accord pillars into product safety systems.

Third Party Partnerships: Furthermore, Microsoft has developed tools and multistakeholder partnerships to combat the rise of misinformation that can pose risks to public health or negatively impact physical or mental wellbeing. Partnering with independent third-party organizations empowers Bing product teams to take additional actions to promote more authoritative information. Microsoft is also a founding member of the C2PA alongside Adobe, Intel, TruePic, Twitter, the BBC, and other tech and media companies. The coalition has an open-source content provenance tool, which allows creators to claim authorship while empowering consumers to make informed decisions about what digital media to trust.

As one example, Microsoft worked with independent third-party partners to perform red team testing of Bing Image Creator to understand the risk of misleading images generated by Bing Image Creator. As part of their analysis, the independent third party prompted Bing Image Creator to create visuals that reinforced or portrayed prominent false narratives related to politics, international affairs, and elections. Based on these evaluations, Microsoft improved existing Bing Image Creator mitigations.

Election Preparation 2024: Bing participated in pre-election discussions and review with CnaM and the European Commission as well as pre-election tabletop exercises prior to the EU Parliament Elections in June 2024 to discuss election-related concerns, highlight Bing election plans, and discuss general election response.

Global Network Initiative Audits: Microsoft also is a member of the GNI, a multistakeholder collaboration to protect freedom of expression and privacy in tech and submits to a regular audit by GNI to ensure Bing has sufficient protections in place to uphold free expression. Bing is biannually audited for its commitments to human rights, including free expression, as part of its membership in the GNI.

Microsoft Bing Ad Library: The Microsoft Bing Ad Library is another initiative aimed at enhancing user understanding and control over the advertising experience. This resource allows users to delve into the specifics of the ads displayed on Bing, including details about the advertisers, their expenditure, and the rationale behind the presentation of certain ads. The "Why This Ad" feature further demystifies the ad selection process by explaining the relevance of ads based on individual online behavior and interests, thereby offering users greater insight and authority over their digital environment.

Policies & Other Mechanisms to Counter Harms

Microsoft Privacy Statement: Users can learn more about Microsoft's collection and use of personal data in the Microsoft Privacy Statement. The Microsoft Privacy Statement outlines what personal data Microsoft collects, how it is used, purposes for which it is used, and how to access and control personal data. This includes how these components apply to specific Microsoft products like search and browse, cookies, Microsoft Account information, and minors' data. Additional topics about personal data are covered such as security, storage, and retention, and how to contact Microsoft.

Responsible Product/Service Design

Ranking Digital Rights Audit: Bing engages in regular self-directed tests and audits of its system to ensure its ability to respond and Microsoft, including Bing, also responds to periodic evaluations by the third-party independent organization Ranking Digital Rights which publishes annual ratings on Bing's practices, governance, and leadership on the protection of freedom of expression and privacy.

Post-Election Reviews: Bing undertakes post-election reviews, as appropriate, to evaluate product and mitigation performance, reflect challenges and learnings, and identify potential areas for improvement. These reviews occur both in product review settings and in broader cross-functional teams dedicated to elections at Microsoft. Bing and Microsoft representatives have also participated in the multi-stakeholder European Commission Elections "tabletop" exercise ahead of the EU Parliamentary Election, related post-election roundtable session, and meetings with CnaM to share practices & reflections.

Metrics x Funnel

Governance

Risk Governance – Roles and Responsibilities: Bing governs and manages safety risks as an essential business function. Bing follows the "Trustworthy Search Principles" to guide the product design, experience, algorithms, and mitigation measures that Bing adopts to ensure users' expectations are met while addressing potential risks or harms arising from use of the service.

- Bing's Corporate Vice President and compliance function, annually and as part of ongoing tracking of business objectives, review how Bing is assessing and managing illegal content and other risks.
- Bing's Digital Safety Program Owner oversees the digital safety practices across Bing products and features, including conducting product design safety reviews, making safety mitigation recommendations, facilitating moderation ops process building, and consulting on measurement and monitoring.
- Bing's content moderation operations function has regular syncs with Bing's digital safety function, which is accountable for policy creation and evaluation, to ensure consistency in the policy understanding and enforcement (e.g., covering work items, edge cases and policy questions, emerging trends)
- Bing's compliance function is responsible for privacy, security and accessibility reviews and building related policies. This function works with respective centralized cross-functional Microsoft teams (privacy, security, and accessibility) responsible for the creation of company-wide policies and operations.
- Bing's customer services and support function is accountable for reviewing and triaging user reports.
- Bing's digital safety function is accountable for search principles creation, evaluation, and operations.
- Bing's legal function provides support to ensure search principles are aligned with legal requirements and central Microsoft policies and standards.
- Bing's Product Managers have weekly deep-dive sessions with cross functional teams within Bing to discuss labelling pipelines, methodology, etc. to ensure alignment and consistency and to identify new risk areas and gaps.

Risk Management

Internal Metrics: Bing routinely reviews the effectiveness of Bing's safety systems through internal metrics as well as social listening channels. Bing maintains a set of metrics to track, monitor and review the efficacy of its interventions on an ongoing basis to ensure that they are performing as expected and not inadvertently introducing additional bias or causing other harm.

Bing uses the DDR as a primary means to measure the efficacy of implemented safety mitigations. DDR measures the likelihood of content violating Bing principles and internal policies appearing in search results or GenAI features across sets of sample queries.

- The metrics are reviewed monthly in shiproom meetings with Bing leaderships and feature teams; when regressions are observed, teams conduct issue identification and implement fixes or drive improvements.
- DDR measurement is also a part of pre-launch assessment.
- Bing conduct analysis on DDR for each DSA systemic risk as an underlying analysis for the formal risk assessment.
- In the areas where Bing observes regressions, Bing conducts issue spotting exercises and further invests in improvements to address the mitigation gaps.

Election Monitoring and Metrics: Bing has improved its monitoring capabilities setting up specific measurement on election-related query sets:

- Bing measures the percentage of responses that are grounded in web results following Bing's authority-driven ranking in Copilot Search. In sample-based testing, Bing has found that 90%+ of Copilot Search responses that were evaluated on a set of elections queries were grounded in web results following Bing's authority-driven ranking.
- Bing monitors jailbreak attempts of Copilot Search.
- Bing also uses social listening to evaluate online conversation around potential responsible AI risks regarding election and political topics and enable mitigation actions.

Bing monitors DDR on an election query set to measure the presence of content that goes against its search policies on different features of Bing Search (including Web, Image, and Video).

- Freshness: Bing measures if up-to-date results are provided and used in grounding responses with factual information on election query set. Given the rapid pace at which elections and campaigns develop, measuring freshness is critical for this risk.
- In addition to metrics, Bing has tested a high number of prompts related to global elections in multiple languages as part of red teaming and applied fixes as needed.

Appendix III: Terms and Acronyms

Below are definitions for the acronyms and initialisms mentioned throughout the report.

Acronym	Definition
1CS	One Compliance Assessment, a platform developed by Microsoft to provide a unified compliance system for engineering groups across the organization.
API	Application Programming Interfaces, a set of rules and protocols that allows different software applications to communicate with each other.
C2PA	Coalition for Content Provenance and Authenticity, an initiative co-founded by Microsoft along with other tech and media companies to develop technical standards for certifying the source and history (or provenance) of media content.
CMP	Content Moderation Platform, an internal tool for implementing content moderation decisions, such as URL exclusion and query treatment.
COPD	Code of Practice on Disinformation, a set of guidelines and standards aimed at combating disinformation and ensuring transparency and accountability in digital platforms.
CnaM	Comisiún na Meán, the media regulator in Ireland, designated Digital Services Coordinator for Ireland under the Digital Services Act.
CSAM	Child Sexual and Abuse Material, refers to all any content that contains or promotes child abuse, or sexual abuse.
CSEAI	Child Sexual Exploitation and Abuse Imagery, refers to visual media that contains or promotes child sexual exploitation or abuse.
DDR	Defensive Defect Rate, primary means to measure the efficacy of implemented safety mitigations and the presence of content that violates Bing policies.
DIRT	Defensive Intelligence Response and Transparency, internal team that is accountable for search principles creation, evaluation, and operations.
DPIA	Data Protection Impact Assessment, required for any new product or feature to investigate possible harms to data subjects.
DSA	Digital Services Act, Article 42.4 (a) and (b) of regulation (EU) 2022/2026
DTSP	Digital Trust & Safety Partnership, an initiative focused on promoting a safer and more trustworthy internet by bringing together leading technology companies to develop industry best practices and assessments for trust and safety in digital services.
EAP	Employee Assistance Program, a program that offers counseling and work/life resources to employees and their household adult family members.
EU	European Union, political and economic union of 27 members states located primarily in Europe.
GDPR	General Data Protection Regulation, a comprehensive data privacy law enacted by the European Union (EU) to enhance individuals' control over their personal data and simplify the regulatory environment for international business by unifying data protection regulations within the EU.

GIFCT	Global Internet Forum to Counter Terrorism, organization, in which Bing has a leadership role, dedicated on combating terrorist, violent, and extremist content.
GNI	Global Network Initiative, a multistakeholder collaboration to protect freedom of expression and privacy in tech.
GPAHE	Global Project Against Hate and Extremism, a nonprofit specializing in tracking and countering hate and extremism globally
IFES	International Foundation for Electoral Systems, an organization focused on strengthening the cybersecurity practices of investigative journalists who are reporting on abuse of state resources in elections.
INCB	United Nations International Narcotics Control Board, an independent and quasi-judicial monitoring body for the production and trade of narcotics and psychotropics, and their availability for medical and scientific purposes.
IP	Intellectual Property, creations of the mind, such as inventions, literary and artistic works, designs, symbols, names, and images used in commerce.
ISD	Institute for Strategic Dialogue, an independent, non-profit organization dedicated to safeguarding human rights and reversing the rising tide of polarization, extremism, and disinformation worldwide.
KPI	Key Performance Indicators, a measurable value that shows how effectively an individual, team, or organization is achieving key business objectives.
MAU	Monthly Active User, a key performance indicator ("KPI") that measures the number of unique users who engage with a product or service within a 30-day period.
MHRA	Medicines and Healthcare products Regulatory Agency, the UK's authority on medicine and device safety.
MSA	Microsoft Services Agreement, Bing's general terms and conditions governing user behavior.
MTAC	Microsoft Threat Analysis Center, a specialized team within Microsoft that focuses on identifying, assessing, and disrupting threats to Microsoft, its customers, and democracies worldwide.
NCII	Nonconsensual Intimate Imagery, also known as "revenge porn."
NDI	National Democratic Institute, an organization aimed at strengthening the cybersecurity support infrastructure for political parties and campaigns internationally.
OFCOM	Office of Communications, regulatory and competition authority for the broadcasting, telecommunications, and postal industries in the United Kingdom.
OKR	Objectives and Key Results, a goal-setting framework used by organizations to define and track objectives and their outcomes.
PAI	Partnership on AI, a non-profit organization that brings together a diverse group of stakeholders, including academic, civil society, industry, and media organizations, to create solutions that ensure artificial intelligence (AI) advances positive outcomes for people and society.
PSA	Public Service Announcement or sometimes also called "Public Safety Announcement", a message in the public interest disseminated without charge, with the objective of raising awareness, changing public attitudes, and behavior towards a social issue.

QC	Quality and Credibility score, a metric used by Bing to evaluate and rank websites. This score helps ensure that users are not misled by problematic material in search results.
RAI	Responsible AI, Microsoft's commitment to cutting-edge research, best-of-breed engineering systems, maintaining a consistently high bar on the policies that Microsoft formulates and promotes, and pursuing excellence with the corporate processes and programs that Microsoft adopts and shares with partners.
SDLC	Software Development Lifecycle, a high-level framework used for developing or changing business applications.
SEO	Search Engine Optimization, the practice of making a public-facing website more visible within search engines, and in turn, to customers and users.
SLA	Service Level Agreement, a contract between a service provider and a customer that defines the types and standards of services to be offered.
StopNCII	Stop Non-Consensual Intimate Image Sharing, a global initiative led by the UK-based Revenge Porn Helpline.
TVEC	Terrorist, Violent, and Extremist Content, refers to content and conduct that can be inherently violent, including calls to action for harm and references to past violent acts.
UNCRC	United Nations Convention on the Rights of the Child, an international human rights treaty that sets out the civil, political, economic, social, health, and cultural rights of children.
VLOP	Very Large Online Platforms, an online platform that reaches an average of 45 million or more active users per month in the European Union (EU) and must comply with the DSA.
VLOSE	Very Large Online Search Engines, an online search engine that reaches an average of 45 million or more active users per month in the EU and must comply with the DSA.

Appendix IV: Summary of External Consultations and Partner Insights

The United Nations International Narcotics Control Board ("INCB") puts efforts to mitigate risks under the DSA related to Illegal Content and Activities and Mental and Physical Wellbeing. Bing engaged with the UN INCB, a treaty-mandated body under the international drug control law. The INCB provided intelligence on drug sales typologies, including key terminology, emojis, and references to dangerous substances and precursor chemicals. These insights were instrumental in improving detection mechanisms for illicit drug-related content and were incorporated into Bing's internal content moderation and alert systems to strengthen service defenses. Further, the insights were added into Bing's internal threat intelligence system (known as the DIRT store), where they are used to support detection, moderation, and safety interventions across the service.

UK Communications Regulator("OFCOM"), to address risks related to Human Dignity, hosted a regulatory workshop where Bing participated in a session focused on online safety for women and girls; with emphasis on combating harassment, intimate image abuse, and gender-based violence. Insights shared during the workshop informed Bing's understanding of vulnerability and risk in gendered online spaces and contributed to internal safety-by-design discussions aimed at better protecting marginalized user groups.

The Lucy Faithfull Foundation, UK (LFF) supports efforts to reduce Illegal Content and Activities. Bing collaborated with the LFF on data exchange related to its CSEAI PSA in the UK. The engagement aimed to assess and improve current user interactions with the PSA while exploring further enhancements, including research on deterrence strategies. This partnership provided valuable input to refine Bing's defensive measures targeting child exploitation content online.

The Medicines and Healthcare products Regulatory Agency, UK ("MHRA") helps to reduce systemic risks associated with health misinformation and illicit pharmaceutical sales. Bing consulted with the MHRA, the UK's authority on medicine and device safety to discuss respective MHRA guidance. Based on this guidance, Bing outlined plans for additional search PSAs, keyword warnings, and other interventions. Some of these actions have already been activated for high-risk search terms, while others remain on the product roadmap pending further coordination. This initiative directly supports compliance with the DSA's mandates on Mental and Physical Wellbeing.

Bing coordinated with Stop Non-Consensual Intimate Image Sharing ("StopNCII") to bolster user protections related to Private and Family Life, particularly in cases involving the non-consensual sharing of intimate images. The DTSP discussion on DSA Reports fulfilled parts of an industry-wide effort to operationalize DSA requirements as Bing took part in a DTSP convening focused on DSA reporting and risk assessment practices. The forum gathered stakeholders from civil society, academia, and major tech companies to exchange methodologies and insights. Participation contributed to the refinement of Bing's systemic risk assessment approach and supported alignment with regulatory expectations on transparency and accountability, addressing all risk areas.

Dr. Dan Reidenberg (Suicide Prevention Expert, U.S.) was consulted in December 2024 on how to improve service responses to users in crisis, particularly regarding self-harm and suicide ("SSI"). Bing engaged with Dr. Dan Reidenberg, a leading expert in suicide prevention to gain further insight and recommendations, of which Bing implemented revised safety banner messaging across several English-speaking markets. These changes enhanced the effectiveness and empathy of Bing's PSAs, ensuring that individuals searching for

sensitive content are met with supportive, clinically informed resources. This initiative falls under the DSA risk area of Mental and Physical Wellbeing.

Bing continued its collaboration with Thorn, an organization committed to defending children from sexual abuse through technology. In alignment with DSA mandates on illegal content, Bing requested updated access to Thorn's CSAM keyword datasets to inform Bing's PSA triggers and moderation lists to remain current and effective. Such engagement can strengthen Bing's ability to detect and disrupt harmful content involving child exploitation across its search and media services.

To further its DSA compliance framework, Bing participated in a joint workshop hosted by the Global Network Initiative ("GNI") and DTSP; which convened industry, civil society, and academic experts to discuss systemic risk assessment under the DSA. This collaboration facilitated a multi-stakeholder exchange of best practices and reinforced Bing's risk identification and evaluation processes across multiple service domains.

To address election-related systemic risks under the DSA, Bing initiated a partnership with the DPA, Germany's national press agency, to receive timely and verified fact-checking data in the run-up to the 2025 elections. This collaboration was designed to inform defensive search interventions targeting information manipulation and manipulative narratives. Implementation began in April 2025, marking a significant step in Bing's civic integrity efforts within the EU regulatory context.

Bing contributed to the Safer Internet Webinar (Child & Youth Safety in AI), which focused on the implications of artificial intelligence for child and youth safety. The session explored how AI technologies shape online experiences for minors, identifying emerging risks such as harmful content amplification, exploitation, and inappropriate personalization. Takeaways from the session informed Bing's trust and safety roadmap, particularly regarding protective interventions in child-oriented search and content environments.

Bing has continued to engage with CnaM through regular meetings throughout the year. During the last year, Bing met with CnaM regarding Article 40's requirements to enable researcher access to data and explained the challenges associated with data minimization and the complexity of data requests in the search context. Bing also responded to a voluntary RFI from CnaM concerning its efforts to support media literacy in Ireland ahead of Ireland's 2024 national elections.

The European Commission held a workshop concerning the DSA systemic risk assessment process and publications requirements for both assessment and audit reports during October 2024 with designated services. The Commission reaffirmed that its only guidance will be in the form of investigations and RFIs. The Commission also noted to the services that redactions in publication must be limited and legally defensible.

Exhibit 299

Message

From: Sunny Xun Liu [redacted@stanford.edu]
Sent: 8/23/2025 12:08:52 AM
To: SMOLINSKA Joanna (EEAS-SAN FRANCISCO) [redacted@eeas.europa.eu]; Jeffrey T. Hancock [redacted@stanford.edu]
CC: Michelle Pualuan [redacted@stanford.edu]
Subject: Re: [request for feedback] research project on automated content moderation

Hi Joanna,

Wed at 10am works great. I will share a zoom calendar with you.

I will reach out to our trust and safety research team to get 3 tickets.

We might have a private roundtable on Sept 24 night about tech regulation compliance and enforcement. I will share more detailed next week if we will have the event. Would love to have you all join us if we host the private event.

Have a good weekend and talk to you next week!

Sunny

Sunny Xun Liu, Ph.D.
Director of Research
Social Media Lab, Department of Communication
Research Scholar
Cyber Policy Center
Stanford
[redacted]@stanford.edu

Produced to HJC

That would be fantastic, thank you, Sunny! I could talk Wed 10-11AM or Thu after 1PM Pacific. Let me know which one is better.

Also, I wanted to kindly enquire about tickets for the T&S conference on Sept 25-26. Could we count on three tickets for the EU Office, for Gerard de Graaf (Head of EU Office in SF), Giulia Geneletti (Policy officer) and me, as EU gov reps?

Many thanks for your help, Sunny, and have a lovely week-end!
Joanna

From: Sunny Xun Liu <[redacted]@stanford.edu>
Sent: 22 August 2025 16:06
To: SMOLINSKA Joanna (EEAS-SAN FRANCISCO) <[redacted]@eeas.europa.eu>; Jeffrey T. Hancock <[redacted]@stanford.edu>
Cc: Michelle Pualuan <[redacted]@stanford.edu>
Subject: Re: [request for feedback] research project on automated content moderation

Hi Joanna,

Great to hear from you!

Jeff is traveling and will be back to office next week.

I love all the questions you shared, if you want to discuss, I have some availability the coming Wed and Thur.

How was your trip back home?

Sunny

Sunny Xun Liu, Ph.D.
Director of Research
Social Media Lab, Department of Communication
Research Scholar
Cyber Policy Center
Stanford
[redacted]@stanford.edu

Dear Jeff, Sunny,

Hoping all is well, I just wanted to check if you had a chance to look at my message below. Would be great to reconnect, let me know if you have any availability next week.

Have a great week-end!
Joanna

Produced to HJC

From: SMOLINSKA Joanna (EEAS-SAN FRANCISCO)
Sent: 17 August 2025 16:51
To: '[redacted]@stanford.edu' <[redacted]@stanford.edu>
Cc: 'Michelle Pualuan' [redacted]@stanford.edu>
Subject: [request for feedback] research project on automated content moderation

Dear Jeff,

I hope this finds you well.

I am reaching out to you in my capacity of a UC Berkeley Tech Policy fellow (n.b. I should be using my UC Berkeley email address for that but I still need to get one). Together with another fellow, we are developing a **survey** to gather expert views on AI-enabled content moderation and dual use content policy (attached). Before we share it more broadly with the T&S community in academia and tech platforms, we wanted to ask for views our trusted contacts. I highly value your expertise and would appreciate your feedback but also understand that the survey is very detailed and might be somewhat daunting, so absolutely no hard feelings if it is too much.

More generally, I was wondering whether you had thoughts on the state of the art of content moderation and where it is going when it comes to automation. More concretely, these are some of the points I'm researching and would love your take on:

[Trends] What main trends do you see shaping the future of AI-powered content moderation systems over the next 1–2 years?

[LLM use] Could you share insights on how large online platforms are integrating LLMs into their content moderation workflows? What are most important limitations and risks?

[Effectiveness] What makes AI-powered content moderation system effective or fit for purpose? What criteria would you use to assess this?

[Governance] What kind of transparency or audit mechanisms would be most effective in evaluating whether AI-driven moderation systems are actually reducing systemic risks on platforms?

[Missing elements] Are these the questions I should be looking at or am I missing something even more important? N.b. I accept that perfect content moderation at scale is unattainable but I see potential in AI for content labelling and moderation. I want to craft a practical tool/ assessment framework to help democratic public authorities monitor use of LLMs, give platforms a reference point backed by experts, and support standard-setting by civil society and researchers.

If you have some time to speak this week it would be fantastic, I am quite flexible on the timing, please let me know.

Take good care!
Joanna

Joanna Smolinska

Counsellor for Digital
Deputy Head of the Office in San Francisco
Delegation of the European Union to the United States
1 Post Street, suite 2300, San Francisco, CA 94 104

Mobile: +1[redacted]
[redacted]@eeas.europa.eu
EU in SF Website

Produced to HJC

Exhibit 300

Stanford | Cyber Policy Center
Freeman Spogli Institute

Compliance and Enforcement in a Rapidly Evolving Landscape

Stanford Dialogue

Where: Stanford University

When: September 24, 2025 at 5:00pm

Organizers and Hosts

- Jeff Hancock, Director, Stanford Cyber Policy Center
- Sunny Xun Liu, Director of Research, Stanford Social Media Lab
- David Evan Harris, Chancellor's Public Scholar, University of California, Berkeley
- KX Jin, Advisory Board Member, Center for Digital Health, Stanford

Purpose & Context

In 2025, the technology sector continues to evolve rapidly, with a need to balance compliance with multi-jurisdictional regulations designed to protect consumers while continuing to drive consumer value.

This roundtable brings together policymakers, academics, and experienced Silicon Valley experts to discuss the state of compliance and enforcement of existing regulations related to online trust and safety. The event will foster open dialogue to identify where data, research, and expertise can enable more effective compliance with and enforcement of existing policy.

Format and Agenda

The event will adopt a roundtable format under Chatham House rule to encourage interactive, high-level discussions rather than formal presentations.

3:00 pm - 5:00 pm: Optional Campus Tour

5:00 pm - 6:00 pm: Dinner and introductions

6:00 pm - 6:10 pm: Welcome, Jeff Hancock

6:10 pm - 6:20 pm: Lightning Keynote: Julie Inman-Grant, Australian eSafety Commissioner

6:20 pm - 7:00 pm: Session 1: Milestones and Pitfalls. Discussion on the state of tech policy compliance and enforcement internationally. Facilitator: David Evan Harris

7:00 pm - 7:10 pm: Break

7:10 pm - 7:50 pm: Session 2: Increasing Regulatory Capacity. Discussion on how academic and tech experts can best support more effective regulatory enforcement. Facilitator: Jeff Hancock

7:50 pm - 8:00pm: Closing Remarks, Jeff Hancock

Expected Outcomes

- A summary report distilling key insights and recommendations, to be shared privately with participants and published in a modified form on Stanford website for broader dissemination.
- Formation of working groups or follow-up initiatives to pilot collaborative evaluation projects.
- Enhanced networks that facilitate future partnerships between regulators, academics and technology experts to support robust development and enforcement of tech policies.

[redacted]

Stanford | Cyber Policy Center
Freeman Spogli Institute

Compliance and Enforcement in a Rapidly Evolving Landscape

Stanford Dialogue

Attendees

Alissa Cooper, Executive Director of the Knight-Georgetown Institute (KGI)

Gerard de Graaf, Senior EU Envoy for Digital to the U.S.

Giulia Geneletti, Policy Officer–Digital Services Act, European Commission and EU Office in San Francisco

Julie Inman Grant, Commissioner, Australia eSafety

Florence G'Sell, Visiting Professor, Stanford Cyber Policy Center

Jeff Hancock, Harry and Norman Chandler Professor of Communication; Director, Stanford Cyber Policy Center

David Evan Harris, Chancellor's Public Scholar, University of California, Berkeley; Professional Faculty, Haas School of Business, UC Berkeley

Ravi Iyer, Managing Director, Neely Center for Ethical Leadership and Decision Making, Managing Director, Psychology of Technology Institute

Vaishnavi J, Principal, Vyanams Strategies

Kang-Xing Jin, Advisory Board Member, Center for Digital Health, Stanford

Tom Kemp, Executive Director, California Privacy Protection Agency

Kathryn King, General Manager of Technology & Strategy, Australia eSafety

Alexios Mantzarlis, Director of the Security, Trust, and Safety Initiative, Cornell Tech

Matt Motyl, Senior Advisor, Psychology of Tech Institute

[redacted]

Mariesa Nicolas, Director of Research, Australia eSafety

Sophie Raim, Program Associate, Project Liberty Institute

Joanna Smolinska, Deputy Head of EU Office in San Francisco

David Sullivan, Founding Executive Director, Digital Trust & Safety Partnership

Kelly Tallon, Executive manager of Industry Compliance and Enforcement, Australia eSafety

Mariana Ferreira Thiele, Deputy Consul of Brazil in San Francisco

Jonathan Porter, Senior Economic Advisor, Ofcom

[redacted]

Exhibit 301

Academic	Former CEO of CrowdTangle	Brandon Silverman
Academic		Eleni Anagnostopoulou
Academic	University of Matej Bel, Slovakia	Jozef Mintál
Academic	George Washington University	Rebekah Tromble
Academic	George Washington University	Anna Lenhart
Academic	Oxford University	Rasmus Nielsen
Academic		Mathias Vermeulen
ACCESS NOW	Access Now	Eliska Pirkova
ACT	Association of Commercial Television	Erard Gilles
ACT	Association of Commercial Television	Greg Polad
ACT	Association of Commercial Television	Sebastiano Bertagni
AER	Association of European Radios	Francesca Fabbri
Algorithm Watch	Algorithm Watch	John Albert
Algorithm Watch	Algorithm Watch	Clara Helming
Carnegie	Carnegie Endowment for International Peace	Samantha Lai
Carnegie	Carnegie Endowment for International Peace	Dean Jackson
CERRE	CERRE	Amaia Ruiz de Gordoa
CERRE	NYU	Laura Edelson
CheckFirst	CheckFirst	Amaury Lesplingarat
CheckFirst	CheckFirst	Nelly Pailleux
D64	D64 - Zentrum für Digitalen Fortschritt	Erik Tuchtfeld
Dentons Global Advisc	Dentons Global Advisors	Julia Grinsnir
EBU	European Broadcasting Union	Julie Lenoir
EBU	European Broadcasting Union	Tadeo Mas Trelles Alvarez
EGTA	Association of Television and Radio Sales Houses	Marie-Pierre Moalic
EGTA	Association of Television and Radio Sales Houses	Conor Murray
EMMA / ENPA	European Magazine Media Association / European New	Jose Guimaras
EPC	European Publishers Council	Angela Mills Wade
EPC	European Publishers Council	Nikolas Moschakis
EPD	European Partnership for Democracy	Lucie Rohrbacherová
EU Disinfo Lab	EU Disinfo Lab	Rita Jonusaite
FakeReporter	FakeReporter	Nitsan Yasur
France Televisions	France Television	Nathalie Gallet
Global Analytics	Global Analytics, Bulgaria	Rumena Filipova
HateAid	HateAid	Pauline Gondesen
IPI	International Press Institute, Austria	Javier Luque Martinez
Liberties	Civil Liberties Union for Europe	Eva Simon
Liberties	Civil Liberties Union for Europe	Franziska Otto
Luminate	Luminate	Swee Leng Harris
NEWS MEDIA	News Media Europe	Aurore Roux
PSSI	Prague Security Studies Institute, Czechia	Kristina Sefcikova
Reuters	Reuters	Stephanie Burnett
SNV	Stiftung Neue Verantwortung	Julian Jaursch
The London Story	Stichting The London Story	Ritumbra Manuvie
The London Story	Stichting The London Story	Alena Kahle
TrustLab	TrustLab	Peter Dudič
Wikimedia	Wikimedia Foundation	Costanza Sciubba Caniglia
ERGA		Andrea Jäger
ERGA		Tove Devries
ERGA		Julian Merer
ERGA		Raphael Honore
ERGA		Julien Giry
ERGA		Ingrid Kools
ERGA		Levente Nyakas
ERGA		Tânia Morais Soares
ERGA		Eulália Pereira
ERGA		Margarida Almeida
ERGA		Myriam Mossong
ERGA		Samra Cindrak
ERGA		Lidia Zebrowska
ERGA		Krzysztofa Jankowskiego

ERGA	Michel Winkels
ERGA	Vanessa Vorteil
ERGA	Michalis Lambrianides
ERGA	Dieter Staudacher
ERGA	Audun Aagre
ERGA	Deborah Molloy Bergin
ERGA	Katarína Drevená
ERGA	Francesco Marrazzo
ERGA	Myriam Mossong
ERGA	Samra Cindrak
ERGA	Lojíková Kateřina
ERGA	Athanasia Vracha
ERGA	Benedict Foll
ERGA	Kjersti Nipen

Exhibit 302

EUROPEAN
COMMISSION

Brussels, 5.12.2025
C(2025) 8630 final

SENSITIVE*

COMMISSION DECISION

of 5.12.2025

pursuant to Articles 73(1), 73(3) and 74(1) of Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act)

Cases DSA.100101, DSA.100102 and DSA.100103 – X (formerly Twitter)

(Only the English text is authentic)

* Distribution only on a 'Need to know' basis - Do not read or carry openly in public places. Must be stored securely and encrypted in storage and transmission. Destroy copies by shredding or secure deletion. Full handling instructions https://europa.eu/!db43PX

TABLE OF CONTENTS

1. Introduction 6
2. Procedure 7
3. The Provider of X in the Union 14
3.1. The Commission's initial position 14
3.2. The arguments of the provider of X 15
3.3. Developments subsequent to the Preliminary Findings 16
3.4. The Commission's identification of the provider of X 17
4. Non-Compliance with Article 25(1) of Regulation (EU) 2022/2065 25
4.1. The legal framework 25
4.2. The relevant measures put in place by the provider of X 25
4.3. The Commission's Preliminary Findings 27
4.3.1. Misappropriating the historical meaning of the verification checkmarks of Twitter on the X service 27
4.3.2. Misappropriating the meaning of cross-industry visual standards 33
4.3.3. Artificial algorithmic amplification in relation to user perception 34
4.3.4. Provider's awareness of the deceptive nature of the 'verified' status 36
4.3.5. Insufficient and ineffective information on meaning of 'verified' status 37
4.3.6. Risk of adversarial abuse of 'verified' status 39
4.3.7. Financial incentives for the provider 41
4.4. The arguments of the provider of X 41
4.5. The Commission's assessment of the provider of X's arguments 43
4.5.1. The claims regarding a lack of legal clarity and guidance 43
4.5.2. The claims regarding alleged general awareness of new meaning of 'verified' status 45
4.5.3. The claims that the Commission's evidence is insufficient 46
4.5.4. The claims questioning the probative value of the survey study 47
4.5.5. The claims regarding alleged shortcomings of Twitter's legacy verification policy 48
4.5.6. The claims regarding alleged inevitability of verification mistakes 48
4.5.7. The claims regarding alleged benefits of X's approach to the 'verified' status 48
4.5.8. The claims regarding alleged imposition of a specific approach to verification 49
4.5.9. The claims regarding alleged adverse incentives of reliance on Risk Assessment Reports 49
4.5.10. The claims regarding alleged exculpatory information and partiality of investigation 50
4.6. Conclusion 50

5. Non-Compliance with Article 39 of Regulation (EU) 2022/2065 51
5.1. The legal framework 51
5.2. The relevant measures put in place by the provider of X 52
5.3. The Commission's Preliminary Findings 56
5.3.1. X's advertising repository is not available through a searchable tool 56
5.3.2. X's advertising repository is not available through a reliable tool 60
5.3.3. X's advertisement repository is not available through API 63
5.4. The arguments of the provider of X 65
5.5. The Commission's assessment of the provider of X's arguments 70
5.5.1. The claim that the Commission has failed to provide guidance 71
5.5.2. The claim that the Commission allegedly misconstrued and misinterpreted main items of evidence on which it relies 71
5.5.3. The claim regarding the conclusions of the independent audit 72
5.5.4. The claim that the search tool is searchable 73
5.5.5. The claim that the search tool is reliable 77
5.5.6. The claim that X's advertising repository is available through an API 81
5.6. Conclusion 82
6. Non-Compliance with Article 40(12) of Regulation (EU) 2022/2065 82
6.1. The legal framework 82
6.2. The relevant measures put in place by the provider of X 83
6.3. The Commission's Preliminary Findings 86
6.3.1. Period between 28 August 2023 and 9 November 2023 86
6.3.2. Period following 9 November 2023 90
6.3.2.1. Overly restrictive interpretation of the eligibility requirements set out in Article 40(12) of Regulation (EU) 2022/2065 91
6.3.2.2. Shortcomings in the application review processes and in communicating outcomes to applicants 96
6.3.2.3. Insufficient data access quotas & durations for Qualified Researchers 98
6.3.2.4. Unnecessary obstacles in the processes set by the provider of X to grant access to data pursuant to Article 40(12) of Regulation (EU) 2022/2065 103
6.3.2.5. Discouraging effect on researchers 105
6.3.3. Prohibiting independent access to data during the entire relevant period 106
6.4. The arguments of the provider of X 108
6.5. The Commission's assessment of the provider of X's arguments 118
6.5.1. The claims that the provider strikes a reasonable balance with its legitimate interests and other applicable rules in the absence of Commission delegated acts and guidance 119

6.5.2. The claims that the provider of X did provide free access to the X API to Qualified Researchers from 28 August 2023 to 9 November 2023 125
6.5.2.1. The claim that the provider of X was transparent regarding the alleged possibility to request access via 'EU-questions@X.com' 125
6.5.2.2. The claim that the absence of a defined process to assess researcher applications and give access to data does not imply that a request would not have been handled correctly 127
6.5.2.3. The claim that no infringement can be inferred from the fact that the provider of X charged researchers a fee for access through its API 127
6.5.3. The claims that the provider of X's interpretation regarding the applicable requirements for data access was justified 129
6.5.3.1. Eligible use of publicly accessible data 129
6.5.3.2. Geographic location of the researchers 136
6.5.3.3. Affiliation to a research organisation 138
6.5.3.4. Disclosure of funding and independence from commercial interests 139
6.5.4. The claims that the provider of X's process to assess applications and grant Qualified Researchers access to data did not suffer from significant shortcomings after 9 November 2023 139
6.5.4.1. The claim that the provider provided all applicants with the reasons for which their application was not accepted, and that the Commission has failed to explain why this explanation is insufficient 139
6.5.4.2. The claim that the provider's processes exceeded what is legally required, including prior to January 2024 140
6.5.4.3. The claim that the provider has swiftly replied to applicants 141
6.5.4.4. The claim that the provider's application form did not create unnecessary friction 142
6.5.4.5. The claim that any comparison with the level of information available to researchers under the Twitter Academic API is irrelevant 142
6.5.5. The claims that the provider of X has provided adequate access quotas and durations to Qualified Researchers after 9 November 2023 143
6.5.5.1. Access quotas 144
6.5.5.2. Access durations 147
6.5.6. The claim that the Commission fails to prove that the practices of the provider discouraged researchers from using their rights pursuant to Article 40(12) of Regulation (EU) 2022/2065 147
6.5.7. The claim that the provider of X does not prohibit scraping to Qualified Researchers, has legitimate reasons not to allow indiscriminate scraping and that the Regulation (EU) 2022/2065 does not prevent providers of very large online platforms from prohibiting scraping 147
6.6. Conclusion 152
7. Alleged procedural errors in the adoption of the Decision 153

7.1. The allegations of the provider of X 153
7.1.1. Access to file and data room process 153
7.1.2. Incompleteness of the Commission's file and its unreliability 154
7.1.3. Breach of rights of defence through the Commission's Decision on the Article 5(6) Implementing Regulation Request. 154
7.2. The Commission's response to the provider of X's allegations 155
7.2.1. Access to file and data room process 155
7.2.2. Completeness and reliability of the Commission's file 161
7.2.3. The Commission's Response to the Article 5(6) Implementing Regulation Request 161
8. Duration of the provider of X's non-compliance 163
8.1. Non-compliance with Article 25(1) of Regulation (EU) 2022/2065 163
8.2. Non-compliance with Article 39 of Regulation (EU) 2022/2065 164
8.3. Non-compliance with Article 40(12) of Regulation (EU) 2022/2065 164
9. Measures to ensure compliance with this Decision 164
9.1. Principles 164
9.2. Application to this case 165
9.2.1. Measures to comply with Article 25(1) of Regulation (EU) 2022/2065 165
9.2.2. Measures to comply with Article 39 of Regulation (EU) 2022/2065 166
9.2.3. Measures to comply with Article 40(12) of Regulation (EU) 2022/2065 166
9.2.4. Enhanced supervision system of remedies to address infringements of obligations laid down in section 5 of Chapter III of Regulation (EU) 2022/2065 167
10. Fines 168
10.1. Principles 168
10.2. Intent or negligence 168
10.2.1. The Commission's Preliminary Findings 168
10.2.2. The arguments of the provider of X 169
10.2.3. The Commission's assessment of the provider of X's arguments 169
10.3. Calculation of the fine 170
10.3.1. Nature of the infringements 170
10.3.1.1. The Commission's Preliminary Findings 170
10.3.1.2. The arguments of the provider of X 170
10.3.1.3. The Commission's assessment of the provider of X's arguments 171
10.3.2. Gravity of the infringements 173
10.3.2.1. The Commission's Preliminary Findings 173
10.3.2.2. The arguments of the provider of X 174

10.3.2.3. The Commission's assessment of the provider of X's arguments 174
10.3.3. Duration of the non-compliance 176
10.3.3.1. The Commission's Preliminary Findings 176
10.3.3.2. The arguments of the provider of X 176
10.3.3.3. The Commission's assessment of the provider of X's arguments 176
10.3.4. Recurrence of the infringements 176
10.3.5. Aggravating circumstances 176
10.3.6. Mitigating circumstances 176
10.3.7. Conclusion 177
10.4. The relevant turnover 177
10.5. Amount of the fine 179
11. Periodic penalty payments 179
11.1. Principles 179
11.2. Application to this case 180

COMMISSION DECISION

of 5.12.2025

pursuant to Articles 73(1), 73(3) and 74(1) of Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act)

Cases DSA.100101, DSA.100102 and DSA.100103 – X (formerly Twitter)

(Only the English text is authentic)

THE EUROPEAN COMMISSION,

Having regard to the Treaty on the Functioning of the European Union,

Having regard to Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act),[1] and in particular Articles 73(1), 73(3) and 74(1) thereof,

Having regard to Commission Decision of 18 December 2023 initiating the proceedings ('the Decision Initiating the Proceedings') and Commission Decision of 12 July 2024 extending the proceedings ('the Decision Extending the Proceedings') adopted pursuant to Article 66(1) of Regulation (EU) 2022/2065 in this case,

Having regard to the Commission Preliminary Findings of 12 July 2024 ('the Preliminary Findings') adopted pursuant to Articles 73(2), 74(3) and 79(1) of Regulation (EU) 2022/2065 in this case,

Having regard to Commission Implementing Regulation (EU) 2023/1201 on detailed arrangements for the conduct of certain proceedings by the Commission pursuant to Regulation (EU) 2022/2065 of the European Parliament and of the Council ('Digital Services Act') ('the Implementing Regulation'), [2] and in particular Article 5 thereof,

Having given the provider of the very large online platform concerned the opportunity of being heard on the Preliminary Findings, including on any matter to which the Commission has taken objections pursuant to Article 79(1) of Regulation (EU) 2022/2065,

After consulting the European Board for Digital Services ('Board') pursuant to Article 66(4) of Regulation (EU) 2022/2065,

Whereas:

1. INTRODUCTION

(1) This Decision adopted pursuant to Articles 73(1), 73(3) and 74(1) of Regulation (EU) 2022/2065 is addressed to X Internet Unlimited Company ('XIUC', formerly Twitter International Unlimited Company 'TIUC'), X Holdings Corp., X.AI Holdings Corp.

[1] OJ L 277, 27.10.2022, pages 1-102.
[2] OJ L 159, 22.6.2023, pages 51-59.

and Mr. Elon Musk (hereinafter collectively referred to as 'the provider of X'). This Decision concerns the provider of X's failure to comply with Articles 25(1), 39 and 40(12) of Regulation (EU) 2022/2065 in relation to the provision of X in the Union.

2. PROCEDURE

(2) By Decision of 25 April 2023,[3] the Commission designated X (formerly Twitter) as a very large online platform pursuant to Article 33(4) of Regulation (EU) 2022/2065 (the 'Designation Decision'). The Designation Decision was addressed to TIUC, which, based on the information available to the Commission at that time, was a subsidiary belonging to the group of legal entities controlled by X Holdings Corp., as the main establishment of the provider of X (formerly Twitter) in the Union within the meaning of Article 56(1) of Regulation (EU) 2022/2065. A copy of the Designation Decision was also sent to X Holdings Corp. and X Corp. Pursuant to Articles 33(6) and 92 of Regulation (EU) 2022/2065, that Regulation started to apply to X as of four months after the notification of the Designation Decision, namely as of 28 August 2023.

(3) Since the Commission had not received a risk assessment report from the provider of X by 28 August 2023 as required under Article 42(4)(a) of Regulation (EU) 2022/2065, on 5 September 2023, the Commission sent that provider a letter containing a set of questions regarding the systemic risks stemming from the design and the functioning of the X service and its related systems in the Union, specifically in relation to their impact on civic discourse and electoral processes.[4] By letter of 20 September 2023, the provider of X replied to those questions.[5] On 14 September 2023, the provider of X submitted to the Commission the risk assessment report for X ('X's Risk Assessment Report').[6]

(4) By letter of 12 October 2023, the Commission addressed to the provider of X a simple request for information ('the first RFI') pursuant to Article 67(1) and (2) and Article 34(3) of Regulation (EU) 2022/2065, regarding that provider's compliance with Article 16, Article 26, Article 27, Article 28(2), Article 34, Article 35(1), Article 38, Article 39, Article 40(12), and Article 41 of that Regulation in relation to the provision of X in the Union.[7] By letters of 20 October 2023[8] and 17 November 2023,[9] the provider of X responded to the first RFI.

(5) On 18 December 2023, the Commission adopted, pursuant to Article 66(1) of Regulation (EU) 2022/2065, the Decision Initiating the Proceedings in cases DSA.100100, DSA.100101, DSA.100102 and DSA.100103 – X (formerly Twitter) with a view to the possible adoption of decisions pursuant to Articles 73 and 74 of

---

3 Commission Decision C(2023) 2721 final.

4 DSA.100101, Doc ID 190; DSA.100102, Doc ID 248; DSA.100103, Doc ID 213.

5 Reply to letter concerning electoral processes (DSA.100101, Doc ID 193-1; DSA.100102, Doc ID 251-1; DSA.100103, Doc ID 216-1).

6 X's Risk Assessment Report (DSA.100101, Doc ID 27-1; DSA.100102, Doc ID 6-1; DSA.100103, Doc ID 9-1).

7 DSA.100101, Doc ID 182-2; DSA.100102, Doc ID 237-2; DSA.100103, Doc ID 194-2.

8 Reply to the first RFI, section I (DSA.100101, Doc ID 1; DSA.100102, Doc ID 1; DSA.100103, Doc ID 1).

9 Reply to the first RFI, sections II–VIII (DSA.100101, Doc ID 2; DSA.100102, Doc ID 2; DSA.100103, Doc ID 2).

Regulation (EU) 2022/2065.[10] By that decision, the Commission initiated proceedings in respect of TIUC's and X Holdings Corp.'s compliance with Article 34(1) and (2), Article 35(1), Article 16(5) and (6), Article 25(1), Article 39, and Article 40(12) of Regulation (EU) 2022/2065. The Decision Initiating the Proceedings was addressed to TIUC, which based on the information available to the Commission at that time was a subsidiary belonging to the group of legal entities controlled by X Holdings Corp., as the main establishment of the provider of X in the Union. A copy of the Decision Initiating the Proceedings was also sent to X Holdings Corp.

(6) On 22 December 2023, the Commission addressed to the provider of X a simple request for information ('the second RFI'), pursuant to Article 67(1) and (2) and Article 34(3) of Regulation (EU) 2022/2065, in cases DSA.100100, DSA.100101, DSA.100102 and DSA.100103 – X (formerly Twitter), regarding that provider's compliance with Article 16(5) and (6), Article 25(1), Article 34(1) and (2), Article 35(1), Article 39, and Article 40(12) of Regulation (EU) 2022/2065 in relation to the provision of X in the Union.[11] By letters of 26 January 2024,[12] 2 February 2024,[13] and 19 February 2024,[14] the provider of X responded to the second RFI.

(7) On 14 March 2024, the Commission addressed to the provider of X a simple request for information ('the third RFI'), pursuant to Article 67(1) and (2) and Article 34(3) of Regulation (EU) 2022/2065, in case DSA.100100 – X (formerly Twitter), regarding that provider's compliance with obligations related to the specific risks of the dissemination of content that was produced by generative artificial intelligence ('AI').[15] By letters of 5 April 2024,[16] and 26 April 2024,[17] the provider of X responded to the third RFI.

(8) By letter of 8 May 2024, the Commission addressed to the provider of X a simple request for information ('the fourth RFI'), pursuant to Article 67(1) and (2) and Article 34(3) of Regulation (EU) 2022/2065, in cases DSA.100100, DSA.100102 and DSA.100103 – X (formerly Twitter), regarding that provider's compliance with Article 16(5) and (6), Article 34(1) and (2), Article 35(1), Article 39, and Article 40(12) of Regulation (EU) 2022/2065 in relation to the provision of X in the Union.[18] By letters of 17 May 2024,[19] 24 May 2024,[20] 27 May 2023,[21] 14 June 2024,[22] and 3 July 2024,[23] the provider of X responded to the fourth RFI.

---

10 Commission Decision C(2023) 9137 final.

11 DSA.100101, Doc ID 40; DSA.100102, Doc ID 11; DSA.100103, Doc ID 12.

12 Reply to the second RFI, section II, request 4; section III, request 5(c); sections V–VII (DSA.100101, Doc ID 164; DSA.100102, Doc ID 210; DSA.100103, Doc ID 163).

13 Reply to the second RFI, section II, request 2, 3; section III, question 4, 5(b), 5(d); section IV (DSA.100101, Doc ID 80; DSA.100102, Doc ID 246; DSA.100103, Doc ID 49).

14 Reply to the second RFI, section I, request 1, 2; section II, request 1; section III, request 1,2,3, 5(a) (DSA.100101, Doc ID 79; DSA.100102, Doc ID 58; DSA.100103, Doc ID 48).

15 DSA.100101, Doc ID 103-2; DSA.100102, Doc ID 137-2; DSA.100103, Doc ID 65-2.

16 Reply to the third RFI, section I (DSA.100101, Doc ID 105; DSA.100102, Doc ID 139; DSA.100103, Doc ID 67).

17 Reply to the third RFI, section II (DSA.100101, Doc ID 107; DSA.100102, Doc ID 141; DSA.100103, Doc ID 69).

18 DSA.100102, Doc ID 144-3; DSA.100103, Doc ID 72-3.

19 Reply to the fourth RFI, section III, requests 4, 12, 16(a)(b)(d); section IV, requests 24–26 (DSA.100101, Doc ID 146; DSA.100102, Doc ID 157; DSA.100103, Doc ID 69).

20 Reply to the fourth RFI, section I, requests 1, 2; section III, requests 16(c)(e)(f) (DSA.100101, Doc ID 148-2; DSA.100102, Doc ID 159-2; DSA.100103, Doc ID 97-2).

(9) By letter of 18 June 2024, the Commission addressed to the provider of X a simple request for information ('the fifth RFI') pursuant to Article 67(1) and (2) of Regulation (EU) 2022/2065 in cases DSA.100100, DSA.100101, DSA.100102 and DSA.100103 – X (formerly Twitter), regarding its corporate structure.[24] By the letter of 20 June 2024, the provider of X responded to the fifth RFI.[25]

(10) In addition to the aforementioned five RFIs addressed to the provider of X, the Commission has gathered information in cases DSA.100101, DSA.100102 and DSA.100103 – X (formerly Twitter) by conducting interviews, pursuant to Article 68(1) and (2) or Regulation (EU) 2022/2065 and by sending requests for information, pursuant to Article 67 of that Regulation, and by carrying out its own testing of the relevant systems of X's online interface. Moreover, on 11 January 2024, 9 April 2024, 23 April 2024, and 29 April 2024, the Commission held meetings with the representatives of the provider of X to discuss the suspected infringements under investigation.

(11) On 12 July 2024, the Commission adopted the Decision Extending the Proceedings to extend the scope of the investigations in cases DSA.100100, DSA.100101, DSA.100102 and DSA.100103 – X (formerly Twitter) to Mr. Elon Musk, who acquired X on 27 October 2022, and to all legal entities directly or indirectly controlled by the latter, as the provider of X in the Union.[26] In the Decision Extending the Proceedings, the Commission defined Mr. Elon Musk together with all the legal entities directly or indirectly controlled by the latter as the 'Musk Group', which does not constitute a registered legal entity in its own right.

(12) On 12 July 2024, the Commission adopted the Preliminary Findings in cases DSA.100101, DSA.100102 and DSA.100103 – X (formerly Twitter), pursuant to Articles 73(2) and 79(1) of Regulation (EU) 2022/2065, in which it took the preliminary view that the provider of X failed to comply with its obligations under Articles 25(1), 39, and 40(12) of Regulation (EU) 2022/2065.[27] In line with the Decision Extending the Proceedings, the Preliminary Findings were addressed to TIUC, as the main establishment of the provider of X in the Union, to X Holdings Corp., as the ultimate legal entity controlling TIUC, and to Mr. Elon Musk, as the natural person ultimately controlling the group of companies to which TIUC and X Holdings Corp. belong.

(13) On 15 July 2024, the Commission notified the Preliminary Findings to the provider of X with a deadline to submit its observations on those Preliminary Findings by 9 August 2024.[28]

---

21 Reply to the fourth RFI, sections VI–VII (DSA.100101, Doc ID 149-2 and 156; DSA.100102, Doc ID 160-2 and 168; DSA.100103, Doc ID 98-2 and 106).

22 Reply to the fourth RFI, section I, requests 1, 2; section II, request 3; section III, request 5(b), 15–22, section V, requests 27–28, section VII (supplementary) (DSA.100101, Doc ID 159 and 194-2; DSA.100102, Doc ID 176 and 252-2; DSA.100103, Doc ID 109 and 217-2).

23 Reply to the fourth RFI, section III, request 5(a), request 6-14, request 23, section V, request 29 (DSA.100101, Doc ID 170-1 and 171-2; DSA.100102, Doc ID 223-1 and 224-2; DSA.100103, Doc ID 178-1 and 179-2).

24 DSA.100101, Doc ID 186-1; DSA.100102, Doc ID 241-1; DSA.100103, Doc ID 198.

25 Reply to the fifth RFI. DSA.100101, Doc ID 189-3; DSA.100102, Doc ID 247-3; DSA.100103, 212-3.

26 Commission Decision C(2024) 5087 final.

27 Commission Decision C(2024) 5087 final.

28 On 12 July 2024, an advance courtesy copy of the Preliminary Findings was provided by the Commission *via* email to TIUC, X Holdings Corp. and Mr. Elon Musk.

(14) On 16 July 2024, the Commission communicated by email a number of organisational points to the provider of X on access to file in the data room, should that provider decide to request access to all documents on the Commission's file, without any redactions, under terms of disclosure that were to be set out in a Commission decision, pursuant to Article 5(3) of Implementing Regulation (EU) 2023/1201.[29]

(15) On 23 July 2024, X Holdings Corp. and TIUC[30] informed the Commission that they had appointed external legal counsel, in relation to the Commission's proceedings covered by the Preliminary Findings. On the same date, the aforementioned external legal counsel sent a letter to the Commission on behalf of X Holdings Corp. and TIUC (the '23 July 2024 Letter') in which they criticised the Commission's procedure for access to the file, while broadly requesting (i) access to copies of non-confidential versions of all documents obtained, produced and/or assembled in the context of the investigation; (ii) access to the file through a physical data room procedure to be limited to what is strictly necessary; (iii) confirmation that the Commission's file is complete and includes minutes of all interviews and meetings between the Commission and third parties; and (iv) clarifications on the deadline to submit observations on the Preliminary Findings.[31]

(16) On 30 July 2024, the Commission replied to the 23 July 2024 Letter. In its reply, the Commission recalled the rules on access to file under Regulation (EU) 2022/2065 and Implementing Regulation (EU) 2023/1201. The Commission also recalled some elements pertaining to the concrete modalities of the exercise of the right to be heard, including on the number of external advisers and obtaining an advance copy of the data room index. The Commission confirmed that the information contained in the Commission's file for the proceedings comprised all documents which were obtained, produced and assembled during the Commission's investigation and that relate to the subject matter of that investigation. Lastly, the Commission provided some clarifications on the deadline for the provider of X to submit observations on the Preliminary Findings.[32]

(17) On 31 July 2024, the external legal counsels sent another letter to the Commission on behalf of X Holdings Corp. and TIUC (the '31 July 2024 Letter') specifically requesting (i) access to copies of non-confidential versions of all documents in the Commission's file outside of the data room, including those mentioned in the Preliminary Findings, pursuant to Implementing Regulation (EU) 2023/1201; (ii) clarifications on the data room procedure; (iii) a confirmation that the Commission's file is complete and includes minutes of all interviews and meetings between the Commission and third parties; and (iv) an extension of the deadline of X Holdings Corp. and TIUC to respond to the Preliminary Findings.[33] On the same date, X Holdings Corp. and TIUC informed the Commission which specified external legal and economic counsel and technical experts should be considered as 'Specified

---

29 DSA.100101, Doc ID 204; DSA.100102, Doc ID 267; DSA.100103, Doc ID 231.

30 The external legal counsel was appointed by TIUC and X Holdings Corp. The power of attorney submitted by the provider of X did not cover Mr. Elon Musk, although the Preliminary Findings were addressed to the latter. (DSA.100101 Doc-ID 209-4; DSA.100102, Doc ID 287-4; DSA.100103, Doc ID 242-9).

31 DSA.100101, Doc ID 209-2, 209-4; DSA.100102, Doc ID 287-2, 287-4; DSA 100103, Doc ID 242-2, 242-4.

32 DSA.100101, Doc ID 209-1, 209-3; DSA.100102, Doc ID 287-1, 287-3; DSA.100103, Doc ID 242-1, 242-3.

33 DSA.100101, Doc ID 210-2; DSA.100102, Doc ID 294-2; DSA.100103, Doc ID 243-2.

External Advisers', in line with Implementing Regulation (EU) 2023/1201, for the purposes of the data room procedure. Although Mr. Elon Musk was identified as one of the addressees of the Preliminary Findings, he did not request access to the Commission's file pursuant to Article 5(1) of Implementing Regulation (EU) 2023/1201.[34]

(18) On 1 August 2024, the Commission replied to the 31 July 2024 Letter.[35] In its reply, the Commission granted X Holdings Corp. and TIUC an extension of the initial deadline to reply to the Preliminary Findings until 5 September 2024. On the same date, X Holdings Corp. and TIUC were granted access to the documents referred to in the Preliminary Findings pursuant to Article 5(2) of Implementing Regulation (EU) 2023/1201. On the same date, the Commission also adopted a decision pursuant to Article 5(3) of Implementing Regulation (EU) 2023/1201 addressed to X Holdings Corp. and TIUC in which it set out the terms of negotiated disclosure for the purpose of access to the Commission's file pursuant to Article 79(4) of Regulation (EU) 2022/2065 and Article 5 of Implementing Regulation (EU) 2023/1201 in cases DSA.100101, DSA.100102 and DSA.100103 (the 'Terms of Disclosure Decision').[36] According to that decision, access to the file pursuant to Article 5(3) and (4) of Implementing Regulation (EU) 2023/1201 in the aforementioned cases was provided to the Specified External Advisers in a data room for a period of five days and under the conditions set out in the data room rules. The list of the Specified External Advisers, the draft acknowledgement of the Terms of Disclosure Decision to be signed by the Specified External Advisers before accessing the data room, and the data room rules (the 'Data Room Rules') were annexed to the Terms of Disclosure Decision.

(19) The data room was accessible on the premises of the Commission's Directorate-General for Communications Networks, Content and Technology (DG CNECT), for five working days, from 5 August to 9 August 2024, between 9:00 and 18:00. The Specified External Advisers decided to access the data room for the entirety of the time available and did not ask for any extension of the duration of the data room access laid down in the Terms of Disclosure Decision. During access to the data room, the Specified External Advisers prepared a data room report as set out in the Terms of Disclosure Decision, Annex C, recitals 12 to 14 (the 'Data Room Report').[37] On 12 August 2024, the Commission shared the Data Room Report with the provider of X, after ensuring that that report did not contain any business secrets or other disaggregated confidential information, in line with the Terms of Disclosure Decision.

(20) On 9 August 2024, the Specified External Advisers sent, on behalf of X Holdings Corp. and TIUC, a letter to the Commission under the procedure foreseen by Article 5(6) of Implementing Regulation (EU) 2023/1201, requesting further access to documents with a view to making such documents directly available to the addressee arguing that the Commission's approach to access to the file presented a number of issues which significantly compromised X Holdings Corp.'s and TIUC's rights of defence (the 'Article 5(6) Implementing Regulation Request'). More specifically, the Specified External Advisers requested further access to non-confidential versions of all documents on the Commission's file with a view to making such documents directly available to X Holdings Corp. and TIUC, as well as their external advisers

34 DSA.100101, Doc ID 210; DSA.100102, Doc ID 294; DSA.100103, Doc ID 243.
35 DSA.100101, Doc ID 221-1; DSA.100102, Doc ID 297-1; DSA.100103, Doc ID 246-1.
36 DSA.100101, Doc ID 234; DSA.100102, Doc ID 309; DSA.100103, Doc ID 259.
37 DSA.100101, Doc ID 233; DSA.100102, Doc ID 308; DSA.100103, Doc ID 258.

outside the data room procedure. As a subsidiary request, the Specified External Advisers requested access, outside the data room procedure, to non-confidential versions of several specific documents consisting of (i) documents provided by a certain national authority; (ii) documents regarding the Commission's communications with a certain third party organisation; (iii) documents provided by a certain third party organisation during a meeting with the Commission; (iv) confidential versions of minutes of meetings with certain researchers; (v) draft intermediate versions of notes of meetings/interviews with researches, the final version of which was cited in the Preliminary Findings; (vi) notes of a meeting and following correspondence with certain researchers, which were not cited in the Preliminary Findings; (vii) the outline of questions asked by the Commission during interviews with researchers in the context of the Commission's investigation in Case DSA.100102; (viii) certain outputs from X's advertising repository, which were not cited in the Preliminary Findings; and (ix) copies of notes of any meetings between the commission and the provider of X in the context of the investigations covered by the Preliminary Findings. In addition, the Specified External Advisers requested (i) an extension of the deadline to respond to the Preliminary Findings, for a period of time equivalent to that lapsing between Article 5(6) Implementing Regulation Request and the date of the Commission's response to that request and (ii) that all of X Holdings Corp.'s and TIUC's external counsel be considered Specified External Advisers even if they did not actually access the data room or had not yet been identified as Specified External Advisers.[38]

(21) On 30 August 2024, and without prejudice to its position on the request made under Article 5(6) of Implementing Regulation (EU) 2023/1201, the Commission decided, pursuant to Article 5(11) of that Regulation, to provide X Holdings Corp. and TIUC with access to 33 documents in the case file of cases DSA.100101, DSA.100102 and DSA.100103, as an interim response to the Article 5(6) Implementing Regulation Request, in order to avoid a disproportionate delay or administrative burden. The aforementioned documents consisted of documents referred to in the following points of recital 20 above: (i) the documents provided by a certain national authority; (vi) final notes of a meeting with certain researchers, which were not cited in the Preliminary Findings; (vii) the outline of the questions asked by the Commission during interviews with researchers in the context of the Commission's investigation in case DSA.100102 and (viii) certain outputs from X's advertising repository, which were not cited in the Preliminary Findings. On the same date, the Commission granted the provider of X an extension of the deadline to reply to the Preliminary Findings until 26 September 2024.[39]

(22) On 20 September 2024, the Commission replied to the Article 5(6) Implementing Regulation Request.[40] In its reply, the Commission rejected the Specified External Advisers' requests under Article 5(6) of the Implementing Regulation (EU) 2023/1201 except for one request, namely that referred to in point (iii) of recital 20 above, which the Commission accepted.

(23) On 26 September 2024, TIUC and X Holdings Corp. submitted their reply to the Preliminary Findings (the 'Reply to the Preliminary Findings').[41] Mr. Elon Musk did not submit any observations to the Preliminary Findings. Since the Preliminary

38 DSA.100101, Doc ID 241; DSA.100102, Doc ID 318-1; DSA.100103, Doc ID 273-1.
39 DSA.100101, Doc ID 246; DSA 100102, Doc ID 323; DSA.100103, Doc ID 278.
40 DSA.100101, Doc-ID 252-1; DSA.100102, Doc-ID 331-1; DSA.100103, 284-1.
41 DSA.100101, Doc ID 233; DSA.100102, Doc ID 337-2; DSA.100103, Doc ID 289-2.

Findings were addressed to TIUC, X Holdings Corp. and Mr Elon Musk as the 'provider of X', this Decision refers to the arguments presented in the Reply to the Preliminary Findings as the arguments of the 'provider of X'.

(24) At a meeting held on 9 August 2024, the Board submitted its views on the Preliminary Findings pursuant to Article 66(4) of Regulation (EU) 2022/2065. The minutes of that meeting were approved without comments or amendments during the following Board meeting of 25 September 2024.[42]

(25) On 23 September 2024, pursuant to Article 42(4)(c) of Regulation (EU) 2022/2065, the provider of X submitted to the Commission its audit report performed under Article 37(4) of Regulation (EU) 2022/2065 (the 'Audit Report 2024').[43]

(26) On 25 February 2025, the provider of X submitted a letter to the Commission requesting '*prompt feedback or guidance in relation to the arguments and questions developed in the Observations submitted on 26 September 2024*'. In that letter, the provider of X also stated that *'[t]o the extent that, as reported by the press, the Commission may believe (wrongly) that there remains a risk of non-compliance with Articles 25(1), 39 and 40(12) DSA in spite of the arguments and evidence put forward in the Defendants' Observations, the Defendants would appreciate a confirmation and an opportunity to discuss this with the Commission*'.[44]

(27) On 27 February 2025, the Commission replied to the letter of 25 February 2025, stating that it will take note of the points raised in that letter and will examine them very carefully. In addition, the Commission stated that its investigation in cases DSA.100101, DSA.100102, and DSA.100103 was still ongoing and that it will continue to assess all relevant aspects in due course. Finally, the Commission stated that it has been and continues to be consistently available for discussions with the provider of X in relation to ongoing cases, should that provider so request.[45]

(28) On 26 March 2025, the provider of X informed the Commission that TIUC would be renamed 'X Internet Unlimited Company' ('XIUC') and that the change was expected to take place on or around 1 April 2025.[46] On 1 May 2025, the provider of X informed the Commission on its new corporate structure referred to in recital 40 below.[47]

(29) On 2 June 2025, the Commission sent a sixth simple request for information ('the sixth RFI') to the provider of X pursuant to Article 67(1) and (2) of Regulation (EU), 2022/2065 to enquire further about the corporate changes referred to in recital 40 below. By letter of 20 June 2025, the provider of X responded to the sixth RFI.[48]

(30) On 1 September 2025, pursuant to Article 42(4)(c) of Regulation (EU) 2022/2065, the provider of X submitted to the Commission its audit report performed under Article 37(4) of Regulation (EU) 2022/2065 (the 'Audit Report 2025').[49]

---

42 European Board for Digital Services, https://digital-strategy.ec.europa.eu/en/policies/dsa-board.
43 DSA.100101, Doc ID 262; DSA.100102, Doc ID 341; DSA.100103, Doc ID 293.
44 DSA.100101, Doc ID 271-1; DSA.100102, Doc ID 371-1; DSA.100103, Doc ID 321-1.
45 DSA.100101, Doc ID 273-2; DSA.100102, Doc ID360-2; DSA.100103, Doc ID 304-2.
46 DSA.100101, Doc ID 282-1; DSA.100102, Doc ID 372-1; DSA.100103, Doc ID322-1.
47 DSA.100101, Doc ID 283; DSA.100102, Doc ID 373; DSA.100103, Doc ID 323.
48 DSA.100101, Doc ID 281; DSA.100102, Doc ID 367; DSA.100103, Doc ID 320.
49 DSA.100101, Doc ID 289-2; DSA.100102, Doc ID 382-2; DSA.100103, Doc ID 325-2.

(31) On 30 September 2025, the provider of X informed the Commission about a banner displayed on X's interface between 18 May 2025 and 23 September 2025 regarding the provider of X's blue check and verification policies.[50]

(32) On 28 November 2025, XIUC's and X Holdings Corp.'s new external legal counsel sent a letter to the Commission requesting access to the file in cases DSA.100101, DSA.100102 and DSA.100103 pursuant to Articles 5(2) and (11) of Implementing Regulation (EU) 2023/1201.[51] The basis for this request was the change in XIUC's and X Holdings Corp.'s legal representation. Additionally, in that same letter the external legal counsel requested feedback on the Reply to the Preliminary Findings and on '*Reinforced Compliance Measures*', such as the one mentioned in recital 31 above. On 4 December 2024, the Commission replied to the letter and rejected the external legal counsel's request.[52]

3. THE PROVIDER OF X IN THE UNION

3.1. The Commission's initial position

(33) In the Decision Extending the Proceedings, the Commission indicated that it considered Mr. Elon Musk and all legal entities directly or indirectly controlled by Mr. Musk, as the provider of X in the Union. That position was based on several elements provided by TIUC in response to the Commission's requests for information.

(34) More specifically, in response to the fourth RFI,[53] TIUC explained that it is a fully owned subsidiary of [redacted]., which is in turn fully owned by [redacted], which is fully owned by X Corp., which is a wholly owned subsidiary of X Holdings Corp. This ownership structure indicates that X Holdings Corp. exercises decisive influence over the conduct of TIUC in the Union. Moreover, in response to the fifth RFI,[54] TIUC explained that Mr. Elon Musk is the majority shareholder of X Holdings Corp., with a [redacted]% stake, while the remainder of X Holdings Corp. is ultimately owned by minority shareholders. In its reply, TIUC also stated that '*Elon Musk ultimately controls X Holdings Corp.*'[55] Finally, TIUC indicated to '*believe*' that Mr. Elon Musk additionally controls Space Exploration Technologies ('Space X'),[56] The Boring Company,[57] Neuralink Corporation ('Neuralink'),[58] and X.AI Corporation ('X.AI').[59]

(35) Based on this information, the Commission took the position that the provider of X in the Union consisted of TIUC, X Holdings Corp., and Mr. Elon Musk, together with all legal entities directly or indirectly controlled by Mr. Elon Musk, namely X Holdings Corp., Space X, The Boring Company, Neuralink Corporation, and X.AI, which the

50 DSA.100101, Doc ID 288-1.
51 DSA.100101, Doc ID 232; DSA.100102, Doc ID 391; DSA.100103, Doc ID 334
52 DSA.100101, Doc ID 233; DSA.100102, Doc ID 392; DSA.100103, Doc ID 335
53 DSA.100101, Doc ID 148-2; DSA.100102, Doc ID 159-2; DSA.100103, Doc ID 97-2.
54 DSA.100101, Doc ID 189-3; DSA.100102, Doc ID 247-3; DSA.100102, Doc ID 212-3.
55 Response to the fifth RFI, question 3(b). DSA.100101, Doc ID 189-3; DSA.100102, Doc ID 247-3; DSA.100103, 212-3.
56 Space X is a spacecraft manufacturer, launch service provider, and defence contractor, and satellite communications company.
57 The Boring Company is a tunnel construction services and equipment company.
58 Neuralink is a neurotechnology company.
59 X.AI is a startup company active in the area of artificial intelligence.

Commission referred to as the 'Musk Group', which does not constitute a registered legal entity in its own right.

3.2. The arguments of the provider of X

(36) In its Reply to the Preliminary Findings, the provider of X put forward three arguments challenging the Commission's initial position on the provider of X in the Union.

(37) In the first place, the provider of X argued that, considering the Musk Group as the provider of X responsible for the alleged infringements of Regulation (EU) 2022/2065 runs counter the Designation Decision, Delegated Regulation (EU) 2023/1127, Commission Implementing Decision C(2023) 8187 final of 27 November 2023 determining the supervisory fee applicable to X for 2023 pursuant to Article 43(3) of Regulation (EU) 2022/2065, and Regulation (EU) 2022/2065 itself. First, the provider of X argued that, according to Article 288 TFEU, a 'decision which specifies those to whom it is addressed shall be binding only on them' and, since the Designation Decision was addressed only to TIUC (now, XIUC), that decision's binding effects apply exclusively to TIUC (now, XIUC). According to the provider of X, entities other than TIUC allegedly belonging to the 'Musk Group' cannot be accused of lack of compliance with the provisions of Regulation (EU) 2022/2065, absent a designation decision addressed to each of them. Second, the provider of X argued that, considering the Musk Group as the provider of X in the Union conflicts with the Delegated Regulation (EU) 2023/1127, since that Regulation defines 'provider of designated service or services' as any provider to whom one or more Commission decisions designating a very large online platform or a very large online search engine under Article 33(4) of Regulation (EU) 2022/2065 is addressed.[60] The provider of X also noted that TIUC (now, XIUC) was the sole addressee of the Designation Decision and of Commission Implementing Decision C(2023) 8187 final, which also determined the supervisory fee payable by the provider of X in light of the financial accounts of TIUC (now, XIUC). Third, the provider of X noted that Regulation (EU) 2022/2065 deliberately uses the term 'provider', rather than 'undertaking', asserting that these concepts are distinct and must be treated as such.[61]

(38) In the second place, the provider of X argued that considering the Musk Group as the provider of X in the Union misinterprets the notion of 'provider of intermediary services' and is at odds with legal and economic reality. First, the provider of X argued that the service provided as part of the X online platform is a 'hosting service' and, therefore, the obligations applicable to X can only be fulfilled by the entity concretely providing the service in the Union. The provider of X also noted that this holds true by looking at the e-Commerce Directive,[62] where no functional approach is promoted in relation to the notion of provider. The concept of 'provider of intermediary services', hence, cannot also encompass entities that are not concretely providing those intermediary services, as it would be the case for the entities included under the 'Musk Group'. Second, the provider of X argued that not only are the entities other than X Holdings Corp. allegedly forming part of the Musk Group active

60 Reply to the Preliminary Findings section 2, page 4 (DSA.100101, Doc ID 233; DSA.100102, Doc ID 337-2; DSA.100103, Doc ID 289-2).

61 Reply to the Preliminary Findings pages 6-7.

62 Reply to the Preliminary Findings page 9.

in markets unrelated to the X service, but also they did not even receive a request for information from the Commission in relation to their shareholding structure.[63]

(39) In the third place, the provider of X argued that, considering the Musk Group as the provider of X is based on a manifest misinterpretation and misapplication of the relevant case law. According to the provider of X, the Commission misinterpreted EU competition law cases when it argues that the conduct of a subsidiary can be imputed to its parent where the parent exercises a decisive influence on it. The provider of X observed that the EU Courts set limits to the application of this doctrine by referring to the judgement in *Sumal* (C-882/19, EU:C:2021:800). According to the provider of X, the Court of Justice ('CJEU') required the identification of a specific link in the economic activity of the entities at issue and the subject matter of the alleged infringements to prevent the possibility that '*a subsidiary within such a group could be held liable for infringements committed in the context of economic activities entirely unconnected to its own activity and in which they were in no way involved, even indirectly*'.[64] On this basis, the provider of X argued that the Commission did not investigate properly if there was a link between the entities being part of the 'Musk Group' nor between these entities and the subject matter of the investigation. The Commission, therefore, cannot rely on the concept of single economic entity to inform its interpretation of the notion of 'provider' in the present case.[65]

3.3. Developments subsequent to the Preliminary Findings

(40) On 1 May 2025, the provider of X informed the Commission of the following organisational changes to its corporate structure:[66]

(a) On March 28, 2025, X.AI Corp., an entity incorporated in Nevada, United States, and X Corp. became wholly owned sister companies under a single ultimate parent entity, X.AI Holdings Corp.

(b) X Corp. will continue to operate independently of X.AI Corp., and X Corp.'s direct corporate parent will not change as a result of the transaction mentioned under point (a) above.

(c) Additionally, X Corp. will remain the parent company of XIUC. The platform through which XIUC provides social networking services will continue to be called 'X'.

(d) XIUC will exercise exclusive decision-making authority over the provision of X in the Union.

(41) On 20 June 2025, the provider of X indicated in its reply to the sixth RFI that:[67]

(a) X.AI Holdings Corp. is the ultimate legal entity controlling (in order) X Holdings Corp., X Corp., and XIUC;

(b) [redacted]

63 Reply to the Preliminary Findings page 10.
64 Reply to the Preliminary Findings, paragraph 47.
65 Reply to the Preliminary Findings, page 10.
66 DSA.100101, Doc ID 283; DSA.100102, Doc ID 373; DSA.100103, Doc ID 323.
67 DSA.100101, Doc ID 281-1; DSA.100102, Doc ID 367-3: DSA.100103, Doc ID 320-1.

(c) [redacted]

3.4. The Commission's identification of the provider of X

(42) While Regulation (EU) 2022/2065 consistently uses the term 'provider' of an intermediary service as the economic operator to whom the obligations stemming from that Regulation are addressed and upon whom fines and periodic penalty payments may be levied for infringements of those obligations, that term is not defined in that Regulation. In accordance with the CJEU's case-law, the meaning and scope of a notion used in an instrument of EU secondary law which is not defined therein, must be determined in accordance with its usual meaning in everyday language, while also taking into account the context in which that notion occurs, the objectives pursued by the instrument in question, and the origins of that notion.[68]

(43) In its usual meaning, the term 'provider' refers, generally, to any person, company or economic entity that provides a particular service. This broad reading of the notion of 'provider' is confirmed by the context in which that notion is used, the objectives pursued by Regulation (EU) 2022/2065, and the origins of that notion.

(44) As regards the context in which that notion is used, Regulation (EU) 2022/2065 regulates a particular type of economic activity, namely the provision of intermediary services defined in Article 3(g) of that Regulation, a subcategory of information society services. The due diligence obligations laid down in Chapter III of Regulation (EU) 2022/2065 are focused on the manner in which a particular activity is exercised, rather than on the characteristics of the entities that exercise it. Provided that an activity consists of the provision of an intermediary service, the entities engaged in the provision of that service should be subject to those obligations, regardless of their legal status or the legal structure of any group to which they may belong.

(45) As regards the objectives pursued by Regulation (EU) 2022/2065, namely contributing to the proper functioning of the internal market and ensuring a safe, predictable and trusted online environment in which the fundamental rights enshrined in the Charter are duly protected, those objectives can only be effectively achieved if the due diligence obligations imposed on intermediary service providers apply to the entity engaged in the economic activity in question, irrespective of its legal status and structure. That is particularly the case where a subsidiary operates an online interface through which an intermediary service is provided, but that subsidiary is unable to determine independently all legal and technical conditions of the provision of that service, and rather implements the strategic decisions taken or the instructions given to it by its parent company. In such situations, ensuring compliance with the obligations laid down by Regulation (EU) 2022/2065 will require decisions to be taken at the level of the parent company.

(46) This interpretation is corroborated by the fact that a failure to comply with the due diligence obligations imposed by Regulation (EU) 2022/2065 must be sanctioned taking into account the '*economic capacity of the provider*' (recitals 86, 101 and 117, Article 51(5) of Regulation (EU) 2022/2065). This consideration is further reflected in Articles 52, 74 and 76 of Regulation (EU) 2022/2065, which refer to the annual

[68] Judgment of 14 July 2022, Porsche Inter Auto and Volkswagen, C-145/20, EU:C:2022:572, paragraph 88 and the case-law cited, and Judgment of 30 April 2024, Trade Express-L and DEVNIA TSIMENT, C-395/22 and C-428/22, EU:C:2024:374, paragraph 65.

worldwide turnover of the provider as the basis for the calculation of fines and penalties for infringements of the obligations imposed on providers under Regulation (EU) 2022/2065. Article 3, point (x), of that Regulation in turn defines the notion of 'turnover' as *the amount derived by an undertaking within the meaning of Article 5(1) of Council Regulation (EC) No 139/2004*.

(47) As regards the origins of the notion of 'provider' used in Regulation (EU) 2022/2065, the Commission's proposal did not use that notion consistently in relation to all categories of services covered by the act but rather used the notion of 'online platform', which was itself defined as a 'provider' of the relevant service (Article 2(h) of the proposal). During the legislative negotiations, most of the provisions were changed to be addressed to 'providers of online platforms' and 'providers of very large online platforms'. However, Article 33(1) of Regulation (EU) 2022/2065 still provides that the obligations laid down in Section 5 of Chapter IV of that Regulation '*apply to online platforms and online search engines which have a number of average monthly active recipients of the service in the Union equal to or higher than 45 million, and which are designated as very large online platforms or very large online search engines pursuant to paragraph 4.*' This confirms that what Regulation (EU) 2022/2065 regulates is the provision of a particular service, without consideration being given to the legal status or structure of the entities providing that service.

(48) Finally, the functional approach to the notion of 'provider', which encompasses every entity engaged in an economic activity, such as the provision of intermediary services, regardless of its legal status and the way in which it is financed, is in line with other areas of Union law.[69] Indeed, Union law recognises that different companies belonging to the same group may form an economic unit and therefore may constitute a provider within the meaning of Regulation (EU) 2022/2065, if the companies concerned do not determine independently their own conduct in the market.

(49) Consequently, the notion of 'provider' within the meaning of Regulation (EU) 2022/2065 should be understood to refer to any economic unit engaged in the provision of intermediary services, irrespective of its legal status or structure, which may consist of several legal entities constituting a single economic unit or even a controlling shareholder and those entities.

(50) Applied to the present case, the Commission observes that, while XIUC is legally a distinct company, X Holdings Corp. indirectly holds 100% of the shares of that company through fully owned subsidiaries (i.e. X Corp, which fully owns [redacted], which fully owns [redacted] which fully owns XIUC). As of 28 March 2025, X.AI Holdings Corp. holds 100% of X Holdings Corp.'s shares.[70] According to the case-law of the CJEU in competition cases, where one company has a 100% shareholding in another company, the former is in a position to exercise decisive influence over the latter and there is a rebuttable presumption that the former does in fact exercise such influence.[71] Moreover, it has been confirmed that the conduct of a subsidiary can also be imputed to its parent where the parent exercises a decisive influence over it, namely where that subsidiary does not decide upon its own conduct

69 See, by analogy, Judgment of 19 February 2002, Wouters and Others, C-309/99, EU:C:2002:98, paragraph 46.

70 Reply to the sixth RFI, question 6, DSA.100101, Doc ID 281-1; DSA.100102, Doc ID 367; DSA.100103, Doc ID 320-1.

71 Judgment of 10 September 2009, Akzo Nobel and Others v Commission, Case C-97/08 P,ECLI:EU:C:2009:536, paragraph 60.

on the market independently, but carries out, in all material respects, the instructions given to it by its parent company.[72]

(51) As regards Mr. Elon Musk, the provider of X indicated that [redacted].[73] The Commission considers this sufficient to grant him effective control over X.AI Holdings Corp., X Holdings Corp. and, thus, XIUC. In addition, the following elements further demonstrate effective control exercised by Mr. Musk over X:

(a) X Holdings Corp.'s Articles of Association, as submitted in reply to the fourth RFI,[74] [redacted]. This has been confirmed again by the provider of X in reply to the sixth RFI[75];

(b) The latest available submission made by Twitter, Inc. to the United States Securities and Exchange Commission in 2022 (the 'SEC Submission'),[76] states that Twitter, Inc. entered into a merger agreement with X Holdings I, Inc., X Holdings II, Inc., a wholly owned subsidiary of X Holdings I, Inc. and with Mr. Elon Musk. The SEC Submission further indicates that X Holdings I, Inc. and X Holdings II, Inc. were affiliates of Mr. Elon Musk. According to the SEC Submission, the merger agreement provided that X Holdings II, Inc. would merge with and into Twitter, Inc., with the latter surviving the merger and becoming a wholly owned subsidiary of X Holdings I, Inc.;

(c) An order issued by the United States District Court for the Northern District of California of 1 November 2024 in case *Parag Agrawal, et al., v. Elon Musk, et al.*,[77] states that '*on April 25, 2022, Twitter and Musk, along with 'Musk's wholly owned entities X Holdings I Inc. and X Holdings II, Inc.,' entered into a merger agreement, under which Musk, acting through his above-referenced entities, 'agreed to buy Twitter.' Thereafter, Musk attempted to 'back out of the deal', Twitter filed a lawsuit 'seeking to force him to proceed with the deal',*

---

72 See, by analogy judgments of 10 September 2009, Akzo Nobel and others v Commission, C-97/08 P, EU:C:2009:536, paragraph 61; of 1 October 2013, Elf Aquitaine v Commission, C-521/09 P, EU:C:2011:620, paragraphs 57 and 63; of 19 July 2012, Alliance One International and Standard Commercial Tobacco v Commission and Commission v Alliance One International and Others, joined cases C-628/10 P and C-14/11 P, EU:C:2012:479, paragraphs 43 and 46; of 8 May 2013, ENI v Commission, C-508/11 P, EU:C:2013:289, paragraph 47; of 16 November 2000, Sotra Kopparbergs Bergslags v Commission, C-286/98 P, EU:C:2000:630, paragraph 29; of 23 January 2014, Evonik Degussa et AlzChem v Commission, T-391/09, EU:T:2014:22, paragraph 77; of 11 July 2013, Commission v Stichting Administratiekantoor Portielje, C-440/11 P, EU:C:2013/514, paragraph 41.

73 Reply to the sixth RFI, question 9, DSA.100101, Doc ID 281-1; DSA.100102, Doc ID 367; DSA.100103, Doc ID 320-1.

74 Reply to the fourth RFI, question 2(a). DSA.100102, Doc ID 160-2; 160-5.

75 Reply to the sixth RFI, question 12, DSA.100101, Doc ID 281-1; DSA.100102, Doc ID 367; DSA.100103, Doc ID 320-1.

76 See https://www.sec.gov/Archives/edgar/data/1418091/000141809122000147/twtr-20220630.htm (last accessed on 17 February 2025). DSA.100101, Doc ID 292-4; DSA.100102, Doc ID 385-4; DSA.100103, Doc ID 328-4.

77 The order is available at this link: https://law.justia.com/cases/federal/district-courts/california/candce/3:2024cv01304/425731/73/ (last accessed on 17 February 2025). DSA.100101, Doc ID 293; DSA.100102, Doc ID 386; DSA.100103, Doc ID 329.

and, 'just days before the trial date,' Musk 'agreed to close the deal on its original terms'. 'The merger transaction closed on October 27, 2022', at which time Musk merged Twitter into X, the latter being a company of which he was, inter alia, the Chairman, Sole Director, and controlling shareholder; he also became the 'Administrator of the Plans.'

(d) Mr. Elon Musk has issued several posts on X suggesting that he exercises control over strategic decisions relating to the provision of the X service in the Union. For example, it was Mr. Musk who announced, on 25 November 2022, that Twitter would re-introduce blue-ribbon verification checkmarks alongside grey checkmarks for accounts of governments and gold checkmarks for accounts of companies (i.e., corporate accounts).[78] Similarly, in an earlier post of 1 November 2022, Mr. Elon Musk announced the introduction of a key change regarding the blue-ribbon verification checkmarks on the X service.[79]

(e) Likewise, on 2 February 2023, Mr. Elon Musk issued a post on Twitter,[80] suggesting personal involvement in the decision to restructure the Twitter API offering.[81] This restructuring included the discontinuation of the Twitter Academic API program.[82]

(f) More generally, in a post of 28 March 2025, Mr. Elon Musk announced the acquisition of X by '@xAI'. The Commission understands – as later confirmed by the provider of X in reply to the sixth RFI – that '@xAI' refers to X.AI Holdings Corp. and X refers to XIUC.[83]

---

78 CNBC, https://www.cnbc.com/2022/11/25/elon-musk-says-twitter-to-launch-verified-service-next-week.html, accessed on 21 June 2024 (DSA.100101, Doc ID 161-190).

79 '*Twitter's current lords & peasants system for who has or doesn't have a blue checkmark is bullshit. Power to the people! Blue for $8/month'X website'*. X website, https://x.com/elonmusk/status/1587498907336118274, accessed on 21 June 2024 (DSA.100101, Doc ID 161-170).

80 '*Yeah, free API is being abused badly right now by bot scammers & opinion manipulators. There's no verification process or cost, so easy to spin up 100k bots to do bad things. Just ~$100/month for API access with ID verification will clean things up greatly.*' X Website, https://x.com/elonmusk/status/1621259936524300289?s=20, accessed on 21 November 2025 (DSA.100101, Doc ID 294-29; DSA.100102, Doc ID 387-29; DSA.100103, Doc ID 330-29).

81 '*Starting February 9, we will no longer support free access to the Twitter API, both v2 and v1.1. A paid basic tier will be available instead'* X Website, https://x.com/XDevelopers/status/1621026986784337922, accessed on 21 November 2025 (DSA.100101, Doc ID 294-33; DSA.100102, Doc ID 387-33; DSA.100103, Doc ID 330-33).

82 Under the Twitter Academic API program, academic researchers could obtain free of charge data access via the Twitter API following the submission of a successful application to the provider of Twitter. Politico, https://www.politico.eu/article/elon-musk-twitter-goes-to-war-with-researchers-api/, accessed on 19 June 2024 (DSA.100102, Doc ID 207-93).
See Nature, https://www.nature.com/articles/d41586-023-00460-z, accessed on 25 June 2024 (DSA.100102, Doc ID 207-42).
CNN Business, https://edition.cnn.com/2023/04/05/tech/academic-researchers-blast-twitter-paywall/index.html, accessed on 19 June 2024, (DSA.100102, Doc ID 207-85).
An overview of the application process for the Twitter Academic API program is available at DEV Community, https://dev.to/suhemparack/a-guide-to-teaching-with-the-twitter-api-v2-3n08, accessed on 19 June 2024 (DSA.100102, Doc ID 207-121).

83 '*@xAI has acquired @X in an all-stock transaction. The combination values xAI at $80 billion and X at $33 billion ($45B less $12B debt). Since its founding two years ago, xAI has rapidly become one of the leading AI labs in the world, building models and data centers at unprecedented speed and scale. X is the digital town square where more than 600M active users go to find the real-time source of ground truth and, in the last two years, has been transformed into one of the most efficient companies in the*

(52) Given Mr. Elon Musk's decisive influence over X.AI Holdings Corp., X Holdings Corp. and, consequently, over XIUC, the Commission concludes that the provider of X in the Union consists of XIUC, X Holdings Corp., X.AI Holdings Corp., and Mr. Elon Musk. None of the arguments put forward by the provider of X in its Reply to the Preliminary Findings call that conclusion into question.

(53) In the first place, the provider of X's claim that considering the Musk Group as the provider of X runs counter to the Designation Decision, Delegated Regulation (EU) 2023/1127, Commission Implementing Decision C(2023) 8187 final, and Regulation (EU) 2022/2065 misunderstands the objectives and purpose of those legal instruments and the scope of the Commission's investigatory powers when adopting the aforementioned decisions under those instruments.

(54) First, the provider of X's claim that TIUC (now, XIUC) is the sole addressee of the Designation Decision and that, therefore, the sole entity bound to comply with Regulation (EU) 2022/2065 disregards the fact that the purpose of the Designation Decision was to designate a particular service as a very large online platform as follows from Article 33(4) of that Regulation. That is why the Commission adopts designation decisions in relation to each of the intermediary services that meet the threshold laid down in Article 33(1) of Regulation (EU) 2022/2065, rather than decisions in relation to each of the providers of those services.[84]

(55) The aforementioned claim also disregards the fact that decisions adopted pursuant to Article 288 TFEU must be addressed to entities and cannot be addressed to single economic units. In this respect, it is important to recall, as observed in the Designation Decision, that the Commission '*lack[ed] the investigative powers prior to designation to obtain the necessary information to identify all legal entities forming the economic unit providing the service to be designated.*'[85] Prior to designation, the Commission may not rely on its power to request information pursuant to Article 67 of Regulation (EU) 2022/2065, which only applies after designation and only in circumstances where the Commission has a suspicion that the provider addressed by such a request has infringed that Regulation. At that stage, the Commission may only rely on Article 24(3) of Regulation (EU) 2022/2065, which only permits it to request the provider of online platform or of online search engine to provide it with information on the number of average monthly active recipients of its service in the Union and information as regards the method used to calculate that number, including explanations and substantiation in respect of the data used. That is why the Commission left open the possibility in the Designation Decision that the provider of X in the Union could be broader than TIUC (now, XIUC), but it did not consider it

---

world, positioning it to deliver scalable future growth. xAI and X's futures are intertwined. Today, we officially take the step to combine the data, models, compute, distribution and talent. This combination will unlock immense potential by blending xAI's advanced AI capability and expertise with X's massive reach. The combined company will deliver smarter, more meaningful experiences to billions of people while staying true to our core mission of seeking truth and advancing knowledge. This will allow us to build a platform that doesn't just reflect the world but actively accelerates human progress. I would like to recognize the hardcore dedication of everyone at xAI and X that has brought us to this point. This is just the beginning. Thank you for your continued partnership and support.' X website, https://x.com/elonmusk/status/1905731750275510312 accessed 19 November 2025 (DSA.100101, Doc ID 292-2; DSA.100102, Doc ID 385-2; DSA.100103Doc ID 328-2).

84 It is in this manner than Regulation (EU) 2022/2065 differs from Regulation (EU) 2022/1925, which requires the Commission to adopt a decision designating a particular undertaking providing core platform services as a 'gatekeeper'.

85 Designation Decision, recital 8.

necessary to resolve that issue at that stage. The same is true for Commission Implementing Decision C(2023) 8187 final, since Delegated Regulation (EU) 2023/1127 gives no power to the Commission to request information from providers of designated services subject to an annual supervisory fee.

(56) The aforementioned claim equally disregards the fact that the Designation Decision was addressed to TIUC (now, XIUC) as the 'main establishment' of the provider of X in the Union within the meaning of Article 56 of Regulation (EU) 2022/2065. While Regulation (EU) 2022/2065 recognises in several recitals and provisions that a provider may have multiple establishments, both within and outside the Union, addressing the Designation Decision to TIUC (now, XIUC) as the 'main establishment' of the provider of X in the Union in accordance with the aforementioned provision, amounted to notifying that Designation Decision to a legal entity forming part of the single economic unit providing the X service in the Union and thus to that single economic unit. Indeed, the economic, organisational and legal links between TIUC and the entities directly and indirectly controlling TIUC mean that those entities were aware of the Designation Decision. In any event, copies of the Designation Decision were also sent to X Holdings Corp. and X Corp.

(57) Second, as regards the provider of X's claim that considering the Musk Group as the provider of X in the Union conflicts with Delegated Regulation (EU) 2023/1127 and the Commission's use of TIUC's financial statements for the calculation of its supervisory fee misunderstands Article 5(2) of that Delegated Regulation and misrepresents the decision the Commission has adopted setting the supervisory fee for X. Article 5(2) of Delegated Regulation (EU) 2023/1127 provides that, where a provider of a designated service has consolidated accounts, the consolidated worldwide profits of the group to which that provider belongs must be considered when determining the maximum overall limit of the supervisory fee. The Musk Group does not have consolidated accounts. That is why TIUC (now, XIUC) submitted the latest audited financial statements for the group level at which its accounts are consolidated, namely, X Corp. and not TIUC as the provider of X erroneously claims. However, this does not imply that X Corp. is the ultimate entity providing X in the Union and that ultimate responsibility for compliance with Regulation (EU) 2022/2065 should stop at that level. Rather, it merely reflects the role of X Corp. as the group entity at whose level TIUC's accounts are consolidated. This is further underscored by the distinction between the calculation of the supervisory fee, which is non-punitive in nature, and the calculation of fines under Regulation (EU) 2022/2065 since its Article 74 of Regulation (EU) 2022/2065 refers to the notion of turnover and links that notion to the concept of an 'undertaking' within the meaning of Regulation (EC) 139/2004 , as explained in recital 46 above.

(58) Third, the provider of X's claim that Regulation (EU) 2022/2065 does not use the notion of 'undertaking' when referring to the addressees of its obligations disregards the fact that that notion is specific to the field of EU competition law, whereas Regulation (EU) 2022/2065 does not concern itself with the competitive impact that the activity of undertakings or their concentrations with other undertakings may have on any specific market and/or on its structure. Regulation (EU) 2022/2065 regulates the provision of intermediary services in the Union. EU legislation regulating the provision of services generally refers to the notion of 'provider' in relation to that economic activity. That being said, that legislation defines that notion in disparate ways, sometimes referring to a natural person or legal entity providing the service and sometimes referring to a single economic unit providing the service. As explained in

recitals 42 to 48 above, in the absence of any definition in Regulation (EU) 2022/2065, the notion of 'provider' should be interpretated in accordance with its usual meaning in everyday language, while also taking into account the context in which that notion occurs, the objectives pursued by the instrument in question, and the origins of that notion.

(59) In any event, in the *Sumal* judgment, to which the provider of X refers, the CJEU held that the decisive condition to bring a follow-on damages action against a subsidiary is the existence of one single economic unit that comprises both the subsidiary and its parent entity, and not whether the subsidiary was an addressee of the Commission's decision.[86] While the judgment concerned subsidiaries' liability for antitrust infringements of their parents in the context of private enforcement, the CJEU clarified that the Commission may choose to fine any legal entity belonging to the same economic unit when at least one legal entity within that economic unit commits an antitrust infringement. The fact that a legal entity is not named as an addressee of such a Commission decision does not allow to argue that it was not part of the same economic unit. Therefore, in the present case, inasmuch as Mr. Elon Musk operates as part of the single economic entity providing X in the Union, alongside X.AI Holdings Corp., X Holdings Corp. and XIUC, they should all be held jointly and severally liable for infringements of Regulation (EU) 2022/2065.

(60) In the third place, the provider of X's claim that interpreting the notion of 'provider' broadly is at odds with legal and economic reality disregards the considerations set out in recital 45 above. Effective enforcement of Regulation (EU) 2022/2065 can only be achieved if the obligations laid down in that Regulation apply to the persons or entities that take the ultimate strategic and business decisions to ensure that the provision of the service in the Union complies with obligations stemming from that Regulation. In many cases, whether a particular service is provided in line with the obligations laid down in Regulation (EU) 2022/2065 will require decisions to be taken on the presentation and operation of the online interface through which that service is provided. As explained in recital 70 of that Regulation, the presentation and operation of an online interface lies at the heart of an online platform (which is a sub-category of an intermediary service) provider's business. Consequently, where an intermediary service is operated by a subsidiary forming part of a larger economic unit with one or more entities exercising decisive influence over the subsidiary actually operating such business, ensuring compliance with the obligations laid down by the Regulation will generally presuppose the existence of such decisive influence also over matters relating to the technology underlying that interface. That Mr. Elon Musk exercises such decisive influence over the provision of X in the Union is demonstrated by the elements listed in recital 51 above. Neither argument put forward by the provider of X in this respect warrants a different conclusion.

(61) First, the provider of X's reference to the definition of 'provider' in Directive 2000/31/EC is misplaced. As a preliminary matter, the definition of a notion used in one piece of EU legislation is not dependent on the definition of the same notion in another piece of EU legislation, unless the former explicitly refers to the latter. In any event, Directive 2000/31/EC gives expression, first and foremost, to the freedom of establishment and the freedom to provide services enshrined in Articles 49 and 56 of the Treaty on the Functioning of the European Union ('TFEU'), which form the legal

[86] Judgment of the Court 6 October 2021, Sumal, S.L. v Mercedes Benz Trucks España S.L., Case C-882/19, ECLI:EU:C:2021:800, paragraph 59.

basis for that directive. Only natural persons that are located in a Member State and legal entities established in a Member State may benefit from those freedoms. Those freedoms do not extend to companies from third countries, which do not have their registered office, central administration, or principal place of business within the Union.[87] In contrast, Regulation (EU) 2022/2065 is based solely on Article 114 TFEU, which allows the EU legislature to regulate the exercise of particular economic activities on the Union market regardless of the legal status of the entities performing those activities or the legal structure of any group to which those entities may belong.

(62) Second, the provider of X's claim that to establish liability there should be a specific link between the economic activity of the entities at issue and the subject matter of the alleged infringements takes the *Sumal* judgment on which it relies out of its proper context. In that judgment, the CJEU applied the 'single economic unit' doctrine to establish 'downward' liability for competition law infringements, allowing subsidiaries to face damages claims for infringements involving their parent companies. Such liability is contingent on two conditions: (i) the subsidiary and parent company must form part of a single economic unit; and (ii) there must be a specific link between the subsidiary's business activities and the subject matter of the alleged competition law infringements. The present case does not concern a follow-on action for damages and the establishment of downward liability of subsidiaries for an infringement of the competition laws by its parent company. Rather, in this Decision, the Commission is required to determine whether an infringement of Regulation (EU) 2022/2065 has occurred and whom to hold accountable for that infringement. Moreover, as part of that determination, the Commission seeks to establish upward liability, in relation to which the EU competition case law has been clear and consistent over time. In particular, the judgment in *Skanska* (C-724/17), rendered the same year as the *Sumal* judgment,[88] confirms that there is no requirement to identify a specific link between the activities of the parent company and the infringement when establishing 'upward liability', i.e. holding parent companies liable for competition law infringements of their subsidiaries.

(63) In light of the foregoing, given Mr. Elon Musk's decisive influence over X.AI Holdings Corp., X Holdings Corp. and, consequently, over XIUC, the Commission concludes that the provider of X in the Union consists of XIUC, X Holdings Corp., X.AI Holdings Corp., and Mr. Elon Musk.

---

87 See recital 36 of Directive 2006/123: "*The concept of 'provider' should cover any natural person who is a national of a Member State or any legal person engaged in a service activity in a Member State, in exercise either of the freedom of establishment or of the free movement of services. The concept of provider should thus not be limited solely to cross-border service provision within the framework of the free movement of services but should also cover cases in which an operator establishes itself in a Member State in order to develop its service activities there. On the other hand, the concept of a provider should not cover the case of branches in a Member State of companies from third countries because, under Article 48 of the Treaty, the freedom of establishment and free movement of services may benefit only companies constituted in accordance with the laws of a Member State and having their registered office, central administration or principal place of business within the Community.*" See also the Explanatory Memorandum to the proposal of the e-Commerce Directive, page 19: *'"service provider" This definition is based on Articles 59 and 60 of the Treaty and their interpretation by the case law of the Court of Justice*'.

88 Judgment of 14 March 2019, Vantaan kaupunki v Skanska Industrial Solutions Oy and Others, Case C-724/17, ECLI:EU:C:2019:204.

4. NON-COMPLIANCE WITH ARTICLE 25(1) OF REGULATION (EU) 2022/2065

4.1. The legal framework

(64) Pursuant to Article 25(1) of Regulation (EU) 2022/2065, '*[p]roviders of online platforms shall not design, organise or operate their online interfaces in a way that deceives or manipulates the recipients of their service or in a way that otherwise materially distorts or impairs the ability of the recipients of their service to make free and informed decisions*.'

(65) Recital 67 of Regulation (EU) 2022/2065 clarifies that *'[d]ark patterns on online interfaces of online platforms are practices that materially distort or impair, either on purpose or in effect, the ability of recipients of the service to make autonomous and informed choices or decisions*.' More concretely, that recital explains that *'[p]roviders of online platforms should […] be prohibited from deceiving or nudging recipients of the service and from distorting or impairing the autonomy, decision-making, or choice of the recipients of the service via the structure, design or functionalities of an online interface or a part thereof. This should include, but not be limited to, exploitative design choices to direct the recipient to actions that benefit the provider of online platforms, but which may not be in the recipients' interests.*'

4.2. The relevant measures put in place by the provider of X

(66) In June 2009, the provider of Twitter launched the 'Verified Account' status to address the phenomenon of false identity or high-profile celebrity impersonations on Twitter, which were perceived as confusing for users and unwelcome by those being impersonated.[89] The Verified Account status, along with the blue-ribbon checkmark badge, was made available to a selection of widely known public figures who were considered to be at risk of impersonation. The purpose of the Verified Account status was to ensure that users could easily see which accounts known to the provider of Twitter actually belonged to, or were actually controlled by, the entity under whose name those accounts were operated. To obtain the Verified Account status, users were required, first, to apply for that status and, next, to undergo a vetting process in which they had to confirm their identity, the authenticity of their account, and their 'notability'. This meant that the provider of Twitter had been in contact with the person or entity the account was representing and verified that that person or entity was behind these accounts. While the content (i.e., the tweets) of such an account might not have been directly written or posted by the person in whose name the account had been set up and operated, the provider of Twitter verified that the account concerned was known to, and approved by, the real person or entity behind the account.[90]

(67) On 28 October 2022, Twitter was acquired by the provider of X. On 9 November 2022, the provider of X began to introduce fundamental changes to Twitter's account verification program, which also affected recipients of its service in the Union. Under a new policy, the Verified Account status became a paid feature widely available to paying subscribers who, in exchange for a monthly fee of USD 7.99, acquired access to various features linked to the 'Twitter Blue' subscription, including the 'verified'

89 X website, https://blog.x.com/official/en_us/a/2009/not-playing-ball.html, accessed on 11 June 2024 (DSA.100101, Doc ID 161-126).

90 Mashable, https://web.archive.org/web/20090615040324/http://mashable.com/2009/06/11/twitter-verified-accounts-2, accessed on 5 June 2024 (DSA.100101, Doc ID 161-30).

status and blue-ribbon checkmark badge displayed next to the user in Twitter's online interface.[91]

(68) This new policy resulted in a reduced level of scrutiny for users receiving the 'verified' status compared to the scrutiny practiced prior to that new policy and led to an initial spike of widely reported high-profile impersonations, including public figures such as Donald Trump, Rudy Giuliani and LeBron James, and companies such as Eli Lilly and Valve.[92] These issues prompted the provider of X to pause its new 'verification' feature on 11 November 2022, only two days after its release.[93]

(69) On 25 November 2022, Mr. Elon Musk announced that Twitter would re-introduce blue-ribbon verification checkmarks alongside grey checkmarks for accounts of governments and gold checkmarks for accounts of companies (i.e., corporate accounts).[94]

(70) On 12 December 2022, the 'Twitter Blue' subscription service was made available again for purchase, with the additional requirements that users subscribing to that service would need an account that was at least 90 days old, not have recently changed their profile information, and have a *'verified phone number'* to sign up.[95] The provider of X also announced that *'subscribers with the blue checkmark will get priority ranking in search, mentions, and replies to help lower the visibility of scams, spam and bots.'*[96]

(71) On 24 March 2023, the provider of X announced that it was discontinuing its legacy verification program completely, meaning that legacy verified users that did not subscribe to and pay for the new service would lose their Verified Account status and the blue checkmark.[97]

---

91 The Verge, https://www.theverge.com/2022/11/9/23448317/elon-musk-twitter-blue-verification-live-ios, accessed on 11 June 2024 (DSA.100101, Doc ID 161-94).

92 The Verge, https://www.theverge.com/2022/11/9/23450289/twitter-impersonators-official-mario-musk-jesus-valve, accessed on 11 June 2024 (DSA.100101, Doc ID 161-98); https://www.bbc.com/news/technology-63599553, accessed on 5 June 2024 (DSA.100101, Doc ID 161-74); Mashable, https://mashable.com/article/disney-junior-fake-account-twitter-verified, accessed on 5 June 2024 (DSA.100101, Doc ID 161-146); Reuters, https://www.reuters.com/technology/companies-wary-twitter-checkmark-policy-fuels-imposter-accounts-2023-05-02/, accessed on 5 June 2024 (DSA.100101, Doc ID 161-150); AP News, https://apnews.com/article/twitter-verification-blue-check-elections-elon-musk-5df36ed183d16ec3bf99446e6827bcdd, accessed on 5 June 2024 (DSA.100101, Doc ID 161-154); Politico, https://www.politico.com/news/2023/05/30/fake-aoc-twitter-account-musk-00099319, accessed on 5 June 2024 (DSA.100101, Doc ID 161-158).

93 CNBC, https://www.cnbc.com/2022/11/11/twitter-blue-subscription-disappears-from-app.html, accessed on 11 June 2024 (DSA.100101, Doc ID 161-102).

94 CNBC, https://www.cnbc.com/2022/11/25/elon-musk-says-twitter-to-launch-verified-service-next-week.html, accessed on 21 June 2024 (DSA.100101, Doc ID 161-190).

95 X website, https://blog.x.com/en_us/topics/product/2022/twitter-blue-update, accessed on 5 June 2024 (DSA.100101, Doc ID 161-90); The Verge, https://www.theverge.com/2022/12/12/23506335/twitter-blue-verified-checkmarks-return-impersonation, accessed on 21 June 2024 (DSA.100101, Doc ID 161-202).

96 X website, https://x.com/premium/status/1602426811656413189, accessed on 21 June 2024 (DSA.100101, Doc ID 161-178).

97 X website, https://x.com/verified/status/1639029459557679104, accessed on 21 June 2024 (DSA.100101, Doc ID 161-182).

(72) On 24 July 2023, 'Twitter' was rebranded as 'X', as a consequence of which the subscription service 'Twitter Blue' became 'X Premium'.[98]

(73) According to early reports from 17 August 2023, following the rebranding, the provider of X was working on an ID verification program[99] to help it enforce its policy against *'misleading and deceptive identities'*.[100] These reports were confirmed on 15 September 2023, with public sources detailing that the provider of X had partnered with the third-party service provider 'Au10tix' for identity verification using government-issued identity documents. The ID verification program was gradually made available globally, with the program being available in the Union at least as early as May 2024 – but it was and remains to-date voluntary for receiving the 'verified' status.[101]

(74) On 27 October 2023, the provider of X launched additional tiers to its subscription services, distinguishing between 'Basic', 'Premium', and 'Premium+' subscriptions. Each of those tiers included a successively larger 'reply prioritization', leading to higher visibility of the replies of subscribers to posts via algorithmic ranking. This means that replies by users with a paid subscription are more likely to be shown at the top of the reply feed under a given post. Only accounts with Premium and Premium+ subscriptions include the blue-ribbon checkmark 'verified' status in the online interface design, as well as the possibility of additional ID verification.[102]

4.3. The Commission's Preliminary Findings

(75) In its Preliminary Findings, the Commission reached the preliminary conclusion that the provider of X failed to comply with Article 25(1) of Regulation (EU) 2022/2065 because the visual and textual design of X's online interface in relation to accounts subscribed to the X Premium and Premium+ services, as well as the manner in which that interface is organised and operated, deceives recipients of the X service into believing that accounts with a 'verified' status on X are meaningfully vetted by that provider, thereby materially distorting or impairing their ability to make free and informed decisions regarding the authenticity and reliability of those accounts. That conclusion was based on the following considerations.

4.3.1. Misappropriating the historical meaning of the verification checkmarks of Twitter on the X service

(76) In the first place, since November 2022, the provider of X has made significant changes to the process for verifying whether users with a 'verified' status are the persons or entities that they claim to be on X, which make that process ineffective in confirming the identity or authenticity of such users as compared to the Verified

98 Financial Times, https://www.ft.com/content/a4330d6b-03cd-41a2-a391-a08d3580b3d1, accessed on 25 June 2024 (DSA.100101, Doc ID 161-38).

99 Engadget, https://www.engadget.com/x-may-soon-add-id-verification-for-preventing-impersonation-190422905.html, accessed on 21 June 2024 (DSA.100101, Doc ID 161-194).

100 X website, https://help.x.com/en/rules-and-policies/x-impersonation-and-deceptive-identities-policy, accessed on 11 June 2024 (DSA.100101, Doc ID 161-118).

101 TechCrunch, https://techcrunch.com/2023/09/15/x-launches-account-verification-based-on-government-id/, accessed on 11 June 2024 (DSA.100101, Doc ID 161-114); X website, https://help.x.com/en/rules-and-policies/verification-policy, accessed on 5 June 2024 (DSA.100101, Doc ID 161-70).

102 TechCrunch, https://techcrunch.com/2023/10/27/x-is-launching-new-premium-and-basic-subscription-tiers/, accessed on 21 June 2024 (DSA.100101, Doc ID 161-186); X website, https://help.x.com/en/using-x/x-premium, accessed on 5 June 2024 (DSA.100101, Doc ID 161-58).

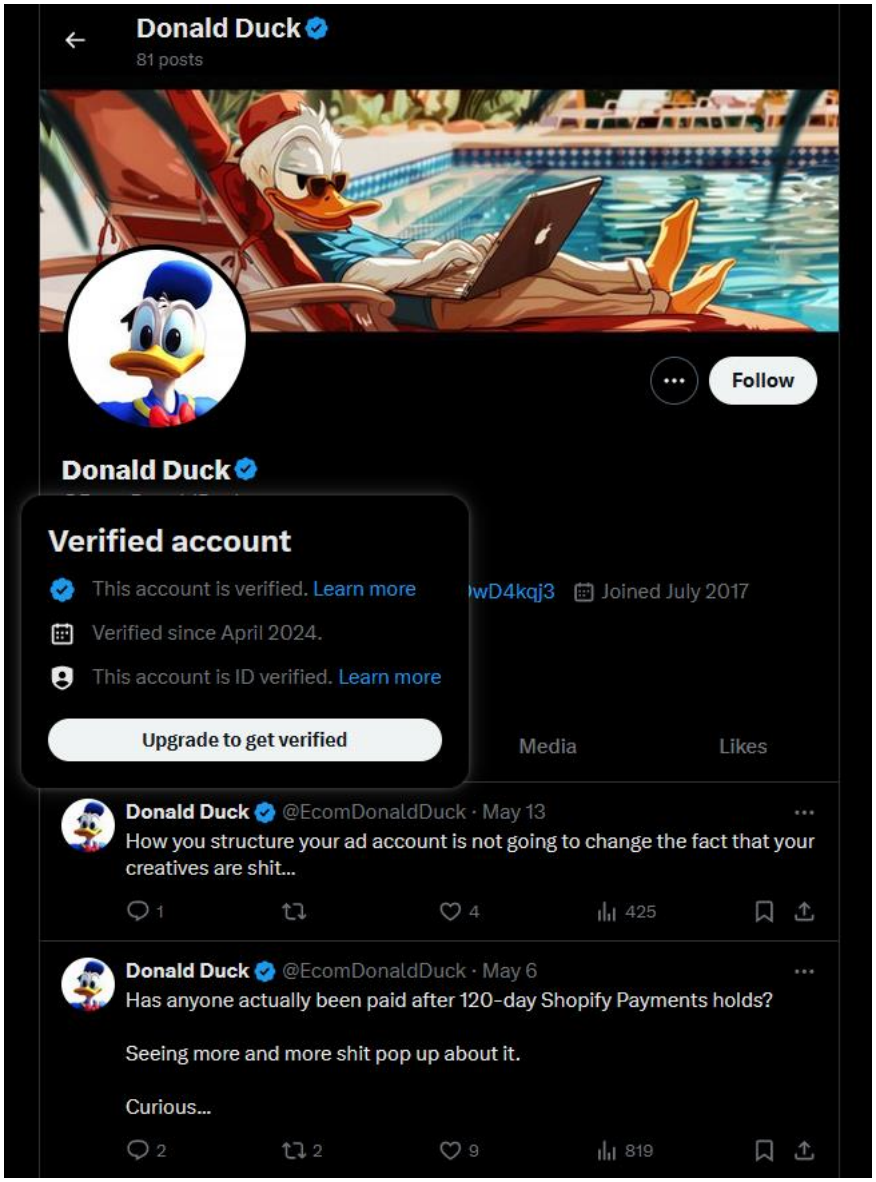
Accounts system that had previously been in place on Twitter, without, however, making any meaningful changes to the design of its online interface in relation to the 'verified' status.

(77) A comparison of the online interface design from 2009 (Figure 1) with the interface design at the time the Preliminary Findings were adopted (Figure 2) in relation to the 'verified' status demonstrates that the latter contains strong visual and textual signifiers that are largely copied from the legacy verification feature displayed in the former.

Figure 1: Original design of blue-ribbon badge with white checkmark from 2009 when Verified Accounts were first launched as a feature.

Source: Mashable, https://web.archive.org/web/20190416180545/https://mashable.com/2009/06/11/twitter-verified-accounts-2/, accessed on 5 June 2024 (DSA.100101, Doc ID 161-30).

Figure 2: Example of the textual and visual online interface design granted to X Premium subscribers. N.B.: The ID verification line is only granted to those who go through an additional verification process.



Source: X website, https://x.com/EcomDonaldDuck, accessed on 21 June 2024 (DSA.100101, Doc ID 161-162).

(78) For users subscribed to X's Premium and Premium+ services, X's online interface displays virtually the same blue-ribbon badge next to the username as that previously available under Twitter's Verified Accounts system – both when displaying the content of such users in the timeline and whenever another user visits that Premium or Premium+ subscribed user's account page.

(79) Moreover, clicking the blue-ribbon badge on the account page of any subscriber to X's Premium and Premium+ services opens a popup window with prominent textual features stating, among others, the following: '*Verified* account', 'This account is *verified*', '*Verified* since [date]', and 'Upgrade to get *verified*' [emphases added]. In its online interface design, the provider of X uses both '*verified*' and '*ID verified'* labels, both marked with the same blue-ribbon badge. Those X Premium and Premium+ subscribers who undergo the additional and voluntary ID verification process receive a line stating that 'This account has been *ID verified*' [emphasis added], the characters 'ID' being the only element allowing to differentiate ID verified accounts from those that not undergone the same process.

(80) However, while the design of X's online interface in relation to the 'verified' status has hardly changed since its introduction in 2009 by the provider of Twitter, the vetting standards and requirements that were previously used to confirm the 'verified' status of users under Twitter's Verified Accounts system have been dismantled and replaced by ineffective standards and requirements under X's Premium and Premium+ subscription services.[103] Instead of the pro-active and thorough up-front vetting process of all applications to attain 'vetted' status under Twitter's previous Verified Accounts system, the provider of X now relies on very limited and probabilistic up-front checks to verify the identity of an account holder when conferring the 'verified' status to subscribers of X's Premium and Premium+ services and, in some instance, on the *ex post* rectification of improperly 'verified' accounts.

(81) According to the provider of X, the system for verifying the authenticity of accounts subscribed to X's Premium and Premium+ services currently in place consists of (i) automated checks of profile pictures and names against existing 'verified' X accounts; and (ii) a manual review whereby '[redacted]' and the '[redacted] *look for signs of impersonation, including for example,* [redacted].' This review process is retriggered whenever a [redacted]'.[104] The provider of X implicitly admits that this initial vetting process cannot provide a clear and unequivocal affirmation of an account's identity or authenticity by simultaneously allowing and encouraging its users to report accounts for violations of its impersonation policy.[105] This means that the provider of X is aware that it cannot rely on its own initial verification process, irrespective of which it grants the 'verified' status to X's Premium and Premium+ services' recipients, including the historically significant blue ribbon checkmark and misleadingly affirmative textual elements such as 'this account is verified'.

---

103 Xiao, Madelyne, Mona Wang, Anunay Kulshrestha, and Jonathan Mayer. 'Account Verification on Social Media: User Perceptions and Paid Enrollment.' arXiv, June 24, 2023. https://doi.org/10.48550/arXiv.2304.14939 (DSA.100101, Doc ID 122-1).

104 Reply to the second RFI, section V, request 1 (DSA.100101, Doc ID 164-9).

105 X website, https://help.x.com/en/safety-and-security/report-x-impersonation, accessed on 11 June 2024 (DSA.100101, Doc ID 161-130).

(82) Furthermore, according to the provider of X, between 28 August 2023 and 26 January 2023, '*X reviewed* [redacted] *new X Premium applications […] In addition, X reviewed* [redacted] *X Premium accounts based on profile changes.*' The provider of X further claims that '*all applications go through human review to a varying degree, including for impersonation review. The reviews are conducted by teams across the* [redacted] *functions comprising* [redacted] *individuals in total (*[redacted] *full-time employees and* [redacted] *contractors).*'[106] Under a conservative assumption that the individuals in charge of this task are in full-time employment and spend between 40 to 60 hours per week reviewing applications, this indicates that the review of each application to X's Premium and Premium+ subscription services took on average no more than 53-79 seconds per account in the period indicated above. Thus, even though the provider of X claims that human review is part of their new 'verification' process, the time allocated to reviewing an account does not appear sufficient to meaningfully claim the 'verified' status for such accounts.

(83) This new 'verification' process is thus a significant departure from the original Verified Accounts system that the provider of Twitter introduced in 2009 to address issues with celebrity impersonations, and which remained largely in place until the November 2022 changes. That system was based on a good-faith effort to thoroughly and proactively verify a user *before* granting the 'verified' status, as opposed to the current limited *ex ante* review coupled with a limited *ex post* sanctioning of impersonator accounts, once such accounts have been identified.[107] Under the previous system, noteworthy users were provided with the label 'Verified Account' and a blue-ribbon badge with a white checkmark (see Figure 1) to indicate that the provider of Twitter had pro-actively confirmed the identity of the real person (or entity, e.g., an undertaking in case of a corporate account) in whose name the account was set up or on whose behalf it was operated. The provider of Twitter had the ability to verify that the account and the content posted on it was known to and approved by that person or entity, including contacting the person or entity the account was representing.[108] While the feature continued to be improved over the years, the underlying pro-active and *ex ante* verification logic did not change until November 2022.[109] In combination with a virtually unchanged overall visual design, this means that, for more than 13 years, users of Twitter could rely on the blue-ribbon badge as a trustworthy indicator that accounts that were vetted by the Verified Accounts program were indeed who they claim to be.

(84) In addition to the reduced thoroughness of the new verification system, the provider of X has also changed the eligibility criteria for affording the 'verified' status from 'active, notable, and authentic' under the previous Twitter's Verified Accounts program to 'complete, active, secure, and non-deceptive' under the X Premium and

106 Reply to the second RFI, section V, request 2 (DSA.100101, Doc ID 164-9).

107 X website, https://blog.x.com/official/en_us/a/2009/not-playing-ball.html, accessed on 11 June 2024 (DSA.100101, Doc ID 161-126).

108 Mashable, https://web.archive.org/web/20190416180545/https://mashable.com/2009/06/11/twitter-verified-accounts-2/, accessed on 5 June 2024 (DSA.100101, Doc ID 161-30).

109 Xiao, Madelyne, Mona Wang, Anunay Kulshrestha, and Jonathan Mayer. 'Account Verification on Social Media: User Perceptions and Paid Enrollment.' arXiv, 24 June 2023. https://doi.org/10.48550/arXiv.2304.14939 (DSA.100101, Doc ID 122-1).

Premium+ subscription services.[110] The change of the 'authenticity' requirement to a 'non-deceptive' requirement marks a departure from the traditional meaning of the verification process. In particular, 'authentic' in the sense of 'representing the entity whose name the account claims' logically restricts the verification process to non-anonymous accounts. That is, an account named, for example, 'Anonymous Joe' could not be verified.[111] The 'non-deceptive' criterion, in contrast, explicitly allows anonymous or fictitious accounts to receive the 'verified' status, while only negatively prohibiting blatant impersonations. This deviates from the previous system – from which the provider of X copied the online interface design – under which recipients of the service had the guarantee that the 'verified' status was reserved for accounts belonging to or controlled by the entity under whose name the account was operated. The Commission considers that, by conferring the 'verified' status including its legacy design and the attached meaning to such anonymous or fictitious accounts, the provider of X lends the credible authenticity assurances of the historical verification program to accounts that cannot be confirmed to be authentic. That provider therefore deceives recipients of its service into believing that accounts with a 'verified' status are meaningfully vetted, thereby materially distorting or impairing their ability to make free and informed decisions regarding the authenticity of those accounts. The aforementioned Figure 2 provides an example of such a fictitious 'verified' user account on X.

(85) Whereas Twitter's Verified Accounts program was designed, organised and operated to provide assurance to recipients of the service that the user they were interacting with had been thoroughly vetted to be who they claimed to be, the 'verified' status for anonymous accounts under X's current verification system does not have the same value or meaning. Based on the current system, that status only means that an account has a verifiable phone number, credit card, and has not been detected by X to be impersonating an actual entity (see recitals 81 and 82 above). Moreover, while the different value of the 'verified' status might be obvious to the users of X for *some* fictitious accounts (e.g., Figure 2), the lenient requirements put in place by the provider of X to obtain that status opens the door to malicious actors – e.g., foreign governments trying to interfere in electoral processes – deceptively posing, for example, as regular citizens to deceive recipients of X's service about the authenticity of their content thereby materially distorting or impairing their ability to make free and informed decisions within the meaning of Article 25(1) of Regulation (EU) 2022/2065.

(86) Consequently, the Commission takes the view that, by combining a lenient vetting system that is significantly less rigorous than that previously in place with the same historical online interface design that used to indicate that the identity and authenticity of an account holder had been 'verified', the provider of X deceives the recipients of the X service into believing that accounts with a 'verified' status on X are meaningfully vetted by that provider thereby materially distorting or impairing their ability to make free and informed decisions regarding the authenticity and reliability of those accounts. In reality, the new vetting process of the provider of X is

---

110 Xiao, Madelyne, Mona Wang, Anunay Kulshrestha, and Jonathan Mayer. 'Account Verification on Social Media: User Perceptions and Paid Enrollment.' arXiv, 24 June 2023. https://doi.org/10.48550/arXiv.2304.14939 (DSA.100101, Doc ID 122-1).

111 While limited exceptions to this might have existed previously due to the vagueness of the 'authenticity' concept, the Commission is not aware of any concrete cases and would consider such cases exceptions to the general rule.

fundamentally different and – for the reasons explained above – far less rigorous than its historical predecessor from which it assimilated the online interface design.

(87) That discrepancy between the historically acquired meaning of the 'verified' status and the new 'verified' status deceives the recipients of the X services as to the authenticity of the accounts that they encounter on X. A study conducted to explore user perceptions and discernment of the new meaning of the 'verified' status on X concludes that '*[t]he results demonstrate a significant mismatch between perceptions and reality: more than half of respondents misunderstand Twitter's blue check verification policies to still require proof of identity […]. Correlation analysis of survey responses suggests that people who are older or have lower digital literacy may be modestly more likely to misunderstand Twitter verification.*'[112]

(88) In other words, the 'verified' status that users subscribing to the X Premium and Premium+ services obtain is deceptive to recipients of X insofar as it suggests, or least does not avoid the suggestion, that the account has been 'verified' according to the same policy and requirements as those underlying Twitter's previous Verified Accounts programme. However, unlike that system, the current system does not provide the guarantee that accounts with the 'verified' status are authentic accounts of persons or entities on whose behalf they are operated, nor can they logically do so in the case of anonymous accounts. Even with the addition of minimal requirements for obtaining the 'verified' status that the provider of X put into place shortly after the initial launch of the X Premium and Premium+ subscription services – notably a confirmed phone number and valid credit card, a certain degree of review against X's impersonation policy, and the voluntary ID verification system put in place in September 2023 – the Commission takes the view that the way that X's online interface is designed, organised and operated is deceptive within the meaning of Article 25(1) of Regulation (EU) 2022/2065.

4.3.2. Misappropriating the meaning of cross-industry visual standards

(89) In the second place, the effective meaning of X's 'verified' status is not only fundamentally and deceptively at odds with the reliability of Twitter's original Verified Accounts program, but also at odds with wider cross-industry standards for similar account verification programmes. Figure 3 depicts various visual markers used by providers of comparable intermediary services to indicate different forms of verification and authenticity assurances. While the eligibility criteria, the precise meaning and purpose of the achieved status, and the linked vetting process might vary across these different services, each of them includes an 'authenticity' criterion and each of them has a generally more rigorous vetting process than that currently employed by the provider of X for the 'verified' status under the X Premium and Premium+ services. For example, while Meta has introduced a paid-for verification scheme for Facebook and Instagram users called 'Meta Verified creator subscriptions', unlike X's 'verified' status, that scheme requires an eligible government issued photo ID to benefit from the verification marker on the online interfaces of Facebook and Instagram.[113] Without prejudice to the question of compliance of those providers with Regulation (EU) 2022/2065, this illustrates that the provider of X is clearly and

[112] Xiao, Madelyne, Mona Wang, Anunay Kulshrestha, and Jonathan Mayer. 'Account Verification on Social Media: User Perceptions and Paid Enrollment.' arXiv, 24 June 2023. https://doi.org/10.48550/arXiv.2304.14939 (DSA.100101, Doc ID 122-1).

[113] Meta website, https://help.instagram.com/2419286908233223?helpref=faq_content, accessed on 13 June 2024 (DSA.100101, Doc ID 161-134).

significantly departing from industry practices regarding according 'verified' status to users.

Figure 3: Screenshot of table from a pre-print paper comparing cross-industry interface designs and eligibility criteria for account verification. 'An asterisk denotes an aspect of a verification program that changed between the survey and publication: Twitter Verified Organizations became available for pay, and Meta both eliminated its 'Notable' and 'Unique' requirements and made verification available for pay. N.B. that 'Twitter Blue' was rebranded to X Premium and Premium+ subscriptions subsequently.

Platform	Program	Launch	Icon	Eligibility Criteria	Open	Requestable	Pay
Twitter	Legacy Verification	2009		"Active," "Notable," and "Authentic"	✗	✓	✗
Twitter	Blue	2022		"Complete," "Active," "Secure," and "Non-deceptive"	✓	✓	✓
Twitter	Verified Organizations	2022		Organizational account, additional criteria ambiguous	✓	✓	✗*
Twitter	Government	2022		Specified account types, additional criteria ambiguous	✓	✓	✗
Facebook	Verified Pages and Profiles	2013		"Notable," "Unique," and "Authentic"*	✓	✓	✗*
Instagram	Verified Badges	2014		"Notable," "Unique," and "Authentic"*	✓	✓	✗*
Snapchat	Snap Stars	2019		"Engagement," "Public Audience," "Authenticity," "Notability," and "Quality"	✓	✗	✗
Snapchat	Public Profiles for Businesses	2021		"Authentic" and "Notable"	✓	✓	✗
TikTok	Verified Accounts	2019		"Active," "Authentic," "Complete," "Notable," and "Secure"	✓	✓	✗
YouTube	Channel Verification	2022		"Authentic," "Complete," and 100,000 subscribers	✓	✓	✗

Source: Xiao, Madelyne, Mona Wang, Anunay Kulshrestha, and Jonathan Mayer. 'Account Verification on Social Media: User Perceptions and Paid Enrollment.' arXiv, June 24, 2023. https://doi.org/10.48550/arXiv.2304.14939 (DSA.100101, Doc ID 122-1).

(90) The study referred to in recital 87 above elaborates that '*major social media platforms have generally followed Twitter's model, offering account verification with similar terminology and check mark iconography […] The primary purpose of account verification – a point of consistency over time and across platforms, with the sole exception of Twitter's recent changes – has been to address account impersonation by affirmatively and proactively confirming an account owner's identity. But verification quickly accumulated other meanings, including as a signifier of account importance and the credibility of content.*'[114] As explained in recitals 81 to 82 above, the provider of X departed from a system of pro-active and *ex ante* confirmation of identity towards a system under which the 'verified' status is distributed to anonymous paying subscribers, with an at least partially *post-hoc* reactive approach to impersonation abuses of the 'verified' status.

(91) The Commission therefore took the view that the design, organisation and operation of X's online interface is deceptive to X's recipients, since the provider of X materially changed the verification process as compared to cross-industry standards, while at the same time maintaining the visual and textual online interface design for 'verified' accounts under Twitter's Verified Accounts program, which was in line with those standards, thereby misappropriating the significance and assurance value of a cross-industry standard for representing accounts whose authenticity and identity has been verified.

4.3.3. Artificial algorithmic amplification in relation to user perception

(92) In the third place, the deceptive nature of the design, organisation and operation of X's online interface is further demonstrated by the combination of misappropriated visual

[114] Xiao, Madelyne, Mona Wang, Anunay Kulshrestha, and Jonathan Mayer. 'Account Verification on Social Media: User Perceptions and Paid Enrollment.' arXiv, June 24, 2023. https://doi.org/10.48550/arXiv.2304.14939 (DSA.100101, Doc ID 122-1).

and textual design elements with an algorithmically amplified visibility of replies from X Premium and Premium+ subscribers. As one of its features, the provider of X offers such subscribers a 'reply prioritization' – i.e., an algorithmically curated higher visibility of their replies to posts on X's online interface – that increases in magnitude as a function of higher tier subscriptions. This prioritization means that replies to other users' posts by accounts having the 'verified' status in X's interface design are statistically more visible in replies to those posts.[115] This effectively allows any subscriber to 'pay for reach' or pay for higher visibility. Crucially, replies from X Premium and Premium+ subscribers enjoy increased visibility without the other recipients of the X service being informed that such higher visibility is due to a financial agreement between the subscriber and the provider of X (as is required in relation to advertising).

(93) For the reasons set out in recitals 94 to 96, the Commission took the view that boosting the visibility of replies from accounts wrongly presented as 'verified' is deceptive to X's recipients, and therefore strengthens the deceptive nature of the design, organisation, and operation of X's online interface regarding the 'verified status.

(94) First, the interface design in the reply-feed of X only shows the historical blue-ribbon badge, without any direct means for users to understand that its significance has changed from the original feature introduced in 2009. To 'Learn more' about the new meaning of this badge, users would need to go to the account page of individual 'verified' users whose post is shown in the reply feed, click on the blue-ribbon icon, and then, in a pop-up window, click on a hyperlink that leads to an explanatory page separate from X's main social media interface.

(95) Second, the selectiveness of the historical and cross-industry applicable standard for acquiring the 'verified' status, as reflected in the still widely applied 'notability' requirement for achieving such a status, means that posts from 'notable' verified accounts have, on average, higher engagement and reach due to the public standing of such accounts. The elimination of the 'notability' requirement follows the proclaimed vision of Mr. Elon Musk in making the 'verified' status accessible to a broader set of users.[116] Under Twitter's original Verified Accounts program, Verified Accounts were organically more visible to users due to their public standing and 'notability'. By artificially boosting content from *any* user that is willing to pay for a subscription, the provider of X *de facto* artificially simulates 'notability' for these users, thereby deceiving the recipients of its service about the significance of their content and therefore materially distorting or impairing their ability to make free and informed decisions regarding those accounts. By treating their content differently in their algorithmic ranking against remuneration, and without disclosing this transparently to the recipients of the X service, the provider of X deceptively suggests that that content is ranked more highly because of its standard recommender system features, such as engagement metrics. This deception impairs the ability of recipients of the X service to

---

115 X website, https://help.x.com/en/using-x/x-premium, accessed on 5 June 2024 (DSA.100101, Doc ID 161-58).

116 See Mr. Elon Musk's post of 1 November 2022: '*Twitter's current lords & peasants system for who has or doesn't have a blue checkmark is bullshit. Power to the people! Blue for $8/month'X website*' – X website, https://x.com/elonmusk/status/1587498907336118274, accessed on 21 June 2024 (DSA.100101, Doc ID 161-170).

make free and informed decisions about the relevance and noteworthiness of the content that they are recommended when visiting the reply-feeds of individual posts.

(96) Third, the deceptive nature of this feature has already been recognised by the French *Commission nationale des comptes de campagne et des financements politique* (CNCCFP). The CNCCFP announced, on 13 July 2023, that, six months before an election, French political candidates would be prohibited from using the visibility boost from X's subscription services. As the CNCCFP explained, this decision was taken because the paid subscription services, which boosts the visibility of an account and its broadcasted messages, are essentially a new form of (undisclosed) sponsored advertising.[117]

4.3.4. Provider's awareness of the deceptive nature of the 'verified' status

(97) In the fourth place, there is evidence demonstrating that the provider of X was and remains aware of the deceptive nature of its online interface design, organisation, and operation for the 'verified' status of X Premium and Premium+ subscribers, but has deliberately chosen to ignore expert advice and has refused to remedy the situation.

(98) Based on publicly available information, Twitter employees conducted an in-depth risk assessment before the November 2022 changes to the verification system resulting in a '*lengthy document warning how the feature could be abused*', which was '*largely ignored*' by the provider of that service.[118] The Commission has interviewed a former Twitter employee who confirmed the content and the relevant aspects of the internal deliberation process. The same employee further explained that the aforementioned document was drafted with the involvement of various teams, including legal and site integrity teams, and was shared with relevant decision-makers before the launch of the 'Twitter Blue' subscription services. In that document, concerns were presented regarding the difference in meaning between the legacy verification program and the new feature, which was liable to confuse users, as well as concerns surrounding impersonations, especially in relation to civic discourse and electoral processes.[119]

(99) These key concerns raised by Twitter employees before the launch of the new subscription services remained valid after Regulation (EU) 2022/2065 started to apply to X and they remain valid to date. After the initial launch of those subscription services, the provider of X implemented certain mitigation measures against impersonation, which is only *one* of the risks identified the internal report mentioned in recital 98 above. The provider of X further explained to the Commission that the risk assessment in that report '*is about an earlier beta version of the X Premium (f.k.a. Twitter Blue) that was launched on November 10, 2022 […]*' and that '*Twitter Blue was relaunched on December 12, 2022 […] with improved mitigation measures in*

---

117 X website, https://x.com/cnccfp_officiel/status/1679494305478483968, accessed on 21 June 2024 (DSA.100101, Doc ID 161-166); https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000023883001, accessed on 13 June 2024 (DSA.100101, Doc ID 161-142); https://www.cnccfp.fr/nos-faq/, point 3.1.1, accessed on 5 June 2023 (DSA.100101, Doc ID 161-78); https://www.euronews.com/next/2023/07/20/french-authorities-ban-twitter-blue-for-political-candidates-ahead-of-elections, accessed on 13 June 2024 (DSA.100101, Doc ID 161-138).

118 NPR, https://www.npr.org/2022/12/02/1140355862/twitters-former-safety-chief-warns-musk-is-moving-fast-and-breaking-things, accessed on 5 June 2024 (DSA.100101, Doc ID 161-82).

119 Non-confidential minutes of interview with former employee of Twitter, 26 January 2024 (DSA.100101, Doc ID 78-2); NPR, https://www.npr.org/2022/12/02/1140355862/twitters-former-safety-chief-warns-musk-is-moving-fast-and-breaking-things, accessed on 5 June 2024 (DSA.100101, Doc ID 161-82).

order to detect and prevent impersonation'.[120] However, as explained in recital 87 above, even with the introduction of these mitigation measures, the concern that the design of X's online interface in relation to the 'verified' status could continue to deceive users in relation to its actual meaning remains unaddressed by the provider of X.

(100) In parallel, in the documents supporting X's Risk Assessment Report, the provider of X acknowledges the risk of user confusion about the online interface design of the 'verified' status for X Premium and Premium+ subscribers in relation to democratic processes, civic discourse, and electoral processes. Although the provider of X claims that account profile labels and checkmarks '*promote authentic experiences and protect against impersonation*' and '*provide our users with confidence that the accounts they are interacting with are authentic and operated by the legitimate owner of an individual or organizational identity*' on its subscription products, it also acknowledges that '*checkmarks that come with subscriptions are intended to be an indicator that an account is real (i.e. that a real human operates an account). Some users may mistakenly interpret this as if the subscriber account's identities have been verified – which is not necessarily the case – and allow such accounts to have undue influence over their civic beliefs*'.[121] As already mentioned in recital 81 above, the provider of X has also implicitly acknowledged that its vetting process for the 'verified' status can never provide a clear and unequivocal affirmation as to the identity or authenticity of users.[122]

(101) Despite these acknowledgements, the provider of X explained to the Commission in response to the second RFI that: '*No research or A/B testing was conducted about user perception of blue checkmarks in relation to the authenticity of the user labelled as 'Verified account' and 'This account is verified' in X (sic) interface design*'.[123] This lack of internal research or testing is notable because the initial internal risk assessment was known to the provider of X and additional independent research on the deceptiveness of the 'verified' status existed prior to the application of Regulation (EU) 2022/2065 to X.[124]

(102) Based on the above, the Commission took the view that the provider of X was well aware of the deceptive nature of its online interface design, organisation, and operation for the 'verified' status of X Premium and Premium+ subscribers, even prior to when Regulation (EU) 2022/2065 started to apply to that service, and that it has taken insufficient action to remedy the situation.

4.3.5. Insufficient and ineffective information on meaning of 'verified' status

(103) In the fifth place, the way the provider of X informs recipients of its service of the value of the 'verified' status is not easily accessible and confusing, thus exacerbating the deceptive design of its online interface, given the historical meaning of the original

120 Reply to the second RFI, section II, request 4 (DSA.100101, Doc ID 164-9).

121 Reply to the first RFI, section VIII, request 9, annex, 'DSA Risk Assessment Actual or Foreseeable Negative Effects on Democratic Processes, Civic Discourse and Electoral Processes – August 2023', page 8 (DSA.100101, Doc ID 2-9).

122 X website, https://help.x.com/en/safety-and-security/report-x-impersonation, accessed on 11 June 2024 (DSA.100101, Doc ID 161-130).

123 Reply to the second RFI, section V, request 3 (DSA.100101, Doc ID 164-9).

124 Xiao, Madelyne, Mona Wang, Anunay Kulshrestha, and Jonathan Mayer. 'Account Verification on Social Media: User Perceptions and Paid Enrollment.' arXiv, June 24, 2023. https://doi.org/10.48550/arXiv.2304.14939 (DSA.100101, Doc ID 122-1).

verification program and the cross-industry standards. The 'Learn more' page, which provides recipients of the X service with that information, is three clicks, a pop-up window, and a separate help page away from the user's timeline experience.

(104) Consequently, the explanations given on the 'Learn more' page do not alleviate the aforementioned concerns regarding the deceptiveness of the 'verified' status but exacerbate them. In the timeline display of the blue-ribbon badge, which is where recipients of the X service spend most of their time, X's recipients are presented with only this historically charged visual 'verification' clue next to the username. Only if and when users visit the actual account indicated with a blue-checked name, can they access more information by clicking first on the blue-ribbon badge on the account page, which opens a popup window as depicted in Figure 2, and then on the 'Learn more' hyperlink displayed next to the textual 'verified' markers. After these three clicks, this hyperlink leads the user to a separate explanatory page detailing the meaning of the blue checkmark and 'verified' status, the policy change compared to the previous verification process, eligibility criteria to receive a blue checkmark, and under which conditions such a checkmark can be revoked.[125]

(105) This explanatory 'Learn more' page on the meaning of the 'verified' status states that the '*blue checkmark means that the account has an active subscription to X Premium and meets our eligibility requirements.*' It further explains that '*Starting April 1, 2023 we began winding down our legacy Verification program and accounts that were verified under the previous criteria (active, notable, and authentic) will not retain a blue checkmark unless they are subscribed to X Premium.*' The current eligibility criteria include the following: '*Complete: Your account must have a display name and profile photo*'; '*Active use: Your account must be active in the past 30 days to subscribe to X Premium*'; '*Security: Your account must have a confirmed phone number*'; '*Non-Deceptive: - Your account must have no recent changes to your profile photo, display name, or username (@handle), - Your account must have no signs of being misleading or deceptive – Your account must have no signs of engaging in platform manipulation and spam*'. Further down, the explanatory page elaborates that '*As a result of this change, X will no longer be accepting applications for the blue Verification checkmarks under the previous criteria (active, notable, and authentic).*'[126]

(106) While the provider of X enumerates its new eligibility criteria (see recital 84 above), it does not explain in meaningful terms that the 'verified' status has an entirely different meaning from what is generally understood and assumed in the industry under the notion of a 'verified' account of a social media platform. Moreover, the aforementioned succinct explanations are difficult to access for recipients of the X service. By leading them three clicks and separate pages away from the main user experience of the timeline, those explanations do not provide meaningful transparency to remedy the initial deception caused by the misappropriation of the legacy online interface design. Finally, the lack of efficiency of the 'Learn more' page is evidenced by the results of the study mentioned at recital 87 above, which shows that more than

[125] X website, https://help.twitter.com/en/managing-your-account/about-x-verified-accounts, accessed on 5 June 2024 (DSA.100101, Doc ID 161-66).

[126] X website, https://help.twitter.com/en/managing-your-account/about-x-verified-accounts, accessed on 5 June 2024 (DSA.100101, Doc ID 161-66).

half of users surveyed do not understand the new requirements and meaning of the 'verified' status.[127]

4.3.6. Risk of adversarial abuse of 'verified' status

(107) In the sixth place, there is evidence that motivated malicious or fraudulent actors benefit from the deceptive nature of the provider of X's 'verified' status for Premium and Premium+ subscribers which materially distorts or impairs the ability of those recipients to make free and informed decisions regarding the authenticity and reliability of those accounts. Exploiting the absence of effective verification standards by the provider of X, these actors can gain the deceptive veneer of authenticity via the 'verified' status to potentially harm recipients of X. For example, a recent analysis of a longitudinal dataset of 2.85 million X accounts, summarised in the study referred to in recital 87 above, shows that there is a significant number of 'verified' accounts boosting cryptocurrency content with potentially deceptive goals. According to that study, '*Twitter accounts […] boosting cryptocurrencies more commonly gain blue check verification after the changes. The cryptocurrency results in particular suggest* ***ongoing strategic exploitation of how people misunderstand verification*** *(emphasis added).*'[128]

(108) In addition, a leading industry expert with long-term experience in disrupting Advances Persistent Threats (APTs) indicated to the Commission that it is unsurprising that abuse of the 'verified' status remains a '*vector for financial crime and financial fraud, for example in relation to crypto scams, impersonations of financial services or businesses providing customer service.*' This is because '*[i]ncreasing the cost of an account on Twitter is one measure that* ***could*** *(emphasis added) be effective against a certain type of spam. It is widely acknowledged amongst experts that some of this activity is financially motivated. So, increasing the unit cost of spam by charging for an account, requiring a phone number, etc – it is not foolproof, but it drives up the costs. Adding an eight-dollar cost to an account does so in a very direct way. This is, however, insufficient to address authenticity questions on platforms like Twitter, because Twitter is a platform targeted by cost-insensitive actors, especially in the context of elections (for instance, there are national governments involved in manipulation campaigns and a $8 amount is trivial for them). Therefore, this does not appear to be a meaningful deterrent against some of the most civically significant forms of manipulation.*'[129]

(109) The view described above is shared and corroborated by a prominent cybersecurity firm that publicly warns that '*[f]raudsters are buying blue checkmarks to impersonate well-known brands on X (ex-Twitter) and scam users.*' It cautions that the '*blue checkmark next to a profile name basically means just one thing: this account has an active X Premium subscription […] the blue checkmark is no longer a guarantee that*

---

127 Xiao, Madelyne, Mona Wang, Anunay Kulshrestha, and Jonathan Mayer. 'Account Verification on Social Media: User Perceptions and Paid Enrollment.' arXiv, June 24, 2023. https://doi.org/10.48550/arXiv.2304.14939 (DSA.100101, Doc ID 122-1).

128 Xiao, Madelyne, Mona Wang, Anunay Kulshrestha, and Jonathan Mayer. 'Account Verification on Social Media: User Perceptions and Paid Enrollment.' arXiv, June 24, 2023. https://doi.org/10.48550/arXiv.2304.14939 (DSA.100101, Doc ID 122-1).

129 Non-confidential minutes of interview with former employee of Twitter, 26 January 2024 (DSA.100101, Doc ID 78-2).

its owner can be trusted. It's just a premium account icon.'[130] These expert views are further corroborated by publicly documented cases of such activity that have regularly appeared on X since the introduction of the 'verified' status feature of the X Premium and Premium+ subscription services and that continue to appear after Regulation (EU) 2022/2065 started to apply to X. In addition to the initial spike of high-profile impersonations mentioned in recital 68 above,[131] there continue to be cases of financially motivated fraudulent abuse of the 'verified' status on X which through the deception distorts or impairs the ability of the recipients of the X service to make free and informed decisions. For example, a customer of a travel booking service complained on X about a missing refund to what they thought was the official account of the travel service provider, but which turned out to be a fake account with a 'verified' status that attempted to defraud users.[132]

(110) More systematically, despite the significantly reduced access to data for external researchers investigating such activities on X (as detailed in section 6 below), there is public evidence of failures of X's verification process to deter APTs and cases in which the deception as to the actual meaning of the 'verified' status feature led to abuse resulting from the distortion or impairment of recipients' ability to make free and informed decisions. For example, an independent researcher from the open-source intelligence ('OSINT') community has published six in-depth analyses of fraudulent activity making use of that feature between May 2023 and May 2024.[133] To give one example from those analyses, a number of 'verified' accounts uploaded what, according to the researcher, appears to be cryptocurrency spam posts in the style of seemingly promoted advertisements and often contained links to a variety of websites related to obscure cryptocurrency and non-fungible token ('NFT') projects of questionable legitimacy or financial value.[134]

---

130 Kaspersky, https://www.kaspersky.com/blog/beware-of-twitter-blue-fake-accounts/49199/, accessed on 5 June 2024 (DSA.100101, Doc ID 161-174).

131 The Verge, https://www.theverge.com/2022/11/9/23450289/twitter-impersonators-official-mario-musk-jesus-valve, accessed on 11 June 2024 (DSA.100101, Doc ID 161-98); BBC, https://www.bbc.com/news/technology-63599553, accessed on 5 June 2024 (DSA.100101, Doc ID 161-74); Mashable, https://mashable.com/article/disney-junior-fake-account-twitter-verified, accessed on 5 June 2024 (DSA.100101, Doc ID 161-146); Reuters, https://www.reuters.com/technology/companies-wary-twitter-checkmark-policy-fuels-imposter-accounts-2023-05-02/, accessed on 5 June 2024 (DSA.100101, Doc ID 161-150); AP News, https://apnews.com/article/twitter-verification-blue-check-elections-elon-musk-5df36ed183d16ec3bf99446e6827bcdd, accessed on 5 June 2024 (DSA.100101, Doc ID 161-154); Politico https://www.politico.com/news/2023/05/30/fake-aoc-twitter-account-musk-00099319, accessed on 5 June 2024 (DSA.100101, Doc ID 161-158).

132 The Guardian, https://www.theguardian.com/technology/2023/aug/27/consumers-complaining-x-targeted-scammers-verification-changes-twitter, accessed on 5 June 2024 (DSA.100101, Doc ID 161-198).

133 Conspirador Norteño Blog, https://conspirator0.substack.com/p/inauthentic-activity-and-twitter, accessed on 5 June 2024 (DSA.100101, Doc ID 161-122); Conspirador Norteño Blog, https://conspirator0.substack.com/p/why-subscription-fees-are-not-a-magic, accessed on 5 June 2024 (DSA.100101, Doc ID 161-34); Conspirador Norteño Blog, https://conspirator0.substack.com/p/blue-checks-ads-and-crypto-spam, accessed on 5 June 2024 (DSA.100101, Doc ID 161-42); Conspirador Norteño Blog, https://conspirator0.substack.com/p/the-nine-thousand-dollar-botnet, accessed on 5 June 2024 (DSA.100101, Doc ID 161-46); Conspirador Norteño Blog, https://conspirator0.substack.com/p/tall-tales-and-fake-faces, accessed on 5 June 2024 (DSA.100101, Doc ID 161-50); Conspirador Norteño Blog, https://conspirator0.substack.com/p/diary-of-a-frosty-fraud, accessed on 5 June 2024 (DSA.100101, Doc ID 161-54).

134 Conspirador Norteño Blog, https://conspirator0.substack.com/p/blue-checks-ads-and-crypto-spam, accessed on 5 June 2024 (DSA.100101, Doc ID 161-42).

(111) The Commission therefore took the view that the evidence set out above further confirmed the deceptive nature of X's online interface design, organisation, and operation with regard to the 'verified' status and that these deceptive interface features impair the ability of recipients of the X service to make free and informed decisions about the authenticity, and in some cases also the security, of the content to which they are exposed on X.

4.3.7. Financial incentives for the provider

(112) In the seventh place, the Commission noted that the provider of X financially benefits from the deceptive nature of that service's online user interface. The provider of X derived revenue from subscriptions to its Premium and Premium+ services and deceptively uses the historical and cross-industry significance of the 'verified' status to drive up subscription numbers. As such, the deceptive design, organisation, and operation of X's online interface directly benefits the provider of X and is not in the interest of the recipients of the X service, insofar as the indiscriminate and deceptive use of the 'verified' status reduces their ability to make free and informed decisions about the authenticity of content on X. Consequently, the design of the interface and practices surrounding the organisation and operation of the 'verified' status additionally fit within the description of 'dark patterns' referred to in recital 67 of Regulation (EU) 2022/2065, which denotes that this '*should include, but not be limited to, exploitative design choices to direct the recipients to actions that benefit the provider of online platforms, but which may not be in the recipients' interests […].*'

4.4. The arguments of the provider of X

(113) In its Reply to the Preliminary Findings, the provider of X challenged the Commission's preliminary findings regarding its non-compliance with Article 25(1) of Regulation (EU) 2022/2065 on the following ten grounds, as set out below.[135]

(114) In the first place, the provider of X argues that there is a lack of legal clarity regarding the scope and meaning of Article 25(1) of Regulation (EU) 2022/2065, and that the Commission has not issued any guidelines on the interpretation of that provision. As set out in Article 25(3) of that Regulation, the Commission may issue guidelines on how Article 25(1) applies to specific practices. The provider of X argues that, instead of providing clearer guidelines or engaging in dialogue with providers, the Commission chose to open proceedings in this case.[136]

(115) Furthermore, referring to the Guidance on the interpretation and application of Directive 2005/29/EC of the European Parliament and of the Council concerning unfair business-to-consumer commercial practices in the internal market (the 'Guidance on Unfair Commercial Practices'), the provider of X argues that the Commission is stretching the notion of 'decisions' '*beyond its natural meaning and intended confines by extending it to textual and visual signifiers that may create a certain user perception, even in circumstances where users are not required to make any decisions or to engage in any behaviour'*. The provider of X argues that there is no precedent or support for the Commission's interpretation of the notion of 'decision' in Article 25(1) of Regulation (EU) 2022/2065, and that this '*would put at risk virtually every online interface implemented by every platform*'.[137] In the second place,

135 Reply to the Preliminary Findings. (DSA.100101, Doc ID 258-2).

136 Reply to the Preliminary Findings, paragraph 83 to 85.

137 Reply to the Preliminary Findings, paragraph 144.

the provider of X claims that the average user of X is aware of the meaning of the 'verified' status due to extensive media coverage on the changed policy, public announcements about this change by the provider of X, and the explanatory webpages where the provider of X elaborates on how the new policy differs from the historical verification program of Twitter. The provider of X argues that it has communicated the changes to its 'verification' and subscription services adequately to the public, as well as the ensuing changes to the meaning of the blue-ribbon checkmark 'verified' status. The provider of X argues that the wide-ranging public attention to these changes support the notion that the average recipient of the X service is aware of the changed meaning of the 'verified' status and that therefore no deception is taking place.[138]

(116) In the third place, the provider of X argues that the evidence put forward by the Commission is insufficient to substantiate a finding of non-compliance with Article 25(1) of Regulation (EU) 2022/2065, and that the Commission has '*failed to assess whether there is a real risk that the average EU X user may be deceived, manipulated or otherwise materially impaired from making any informed decisions*'.[139]

(117) The provider of X alleges that the '*Commission has not conducted any testing or meaningful examination of whether the design of X's online interface is liable to materially impair the ability of average EU X users to make free and informed decisions.*' The provider of X claims that case-law in other areas suggest that the assessment of deceptiveness '*must be conducted against the benchmark of an average consumer that is reasonably critical, conscious and circumspect*' and that the Commission did not conduct such an assessment.[140]

(118) In the fourth place, the provider of X questions the probative value of the survey study on 'Account Verification on Social Media' referred to in recital 87 above in support of the Commission's preliminary findings.[141] The provider of X claims that that study '*was written by three students together with their supervisor*' and '*has not been peer-reviewed nor published in any academic journal*'; is based on a small US-based sample; and was carried out before Regulation (EU) 2022/2065 became applicable and close to the initial release of the new 'verification' program. Substantively, the provider of X claims that the study also shows that a significant number of users were aware of the changed requirements for the 'verified' status of the new program. Furthermore, the provider of X claims that the Commission ignored an allegedly exculpatory study from that paper's bibliography that investigated the links between the perceived credibility of accounts and verification checkmarks.[142]

(119) In the fifth place, the provider of X argues that Twitter's legacy verification policy has always been subject to controversy and points out a November 2017 incident where the provider of Twitter paused the verification program because it caused confusion. The provider of X furthermore points towards an attempt by the provider of Twitter in

---

138 Reply to the Preliminary Findings, paragraphs 102-103, see also paragraphs 94, 96-101, 110-111, 113, 115-140.

139 Reply to the Preliminary Findings, paragraphs 87-88.

140 Reply to the Preliminary Findings, paragraphs 145-146.

141 Xiao, Madelyne, Mona Wang, Anunay Kulshrestha, and Jonathan Mayer. 'Account Verification on Social Media: User Perceptions and Paid Enrollment.' arXiv, June 24, 2023. https://doi.org/10.48550/arXiv.2304.14939 (DSA.100101, Doc ID 122-1).

142 Reply to the Preliminary Findings, paragraphs 147-149.

May 2021 to expand the verification program to the general user base, which was halted due to high demand and some cases of verifying fake accounts.[143]

(120) In the sixth place, the provider of X argues that '*any verification system may have imperfections or be exposed to potential abuse by certain users*' and that it has put mitigation measures in place to reduce these risks. The provider of X claims that these efforts '*are inconsistent with any alleged intention to deceive users*'. Furthermore, in relation to abuses by X users, the provider of X stresses that '*X shall not be subject to general monitoring or active fact-finding obligations*' in accordance with Article 8 of Regulation (EU) 2022/2065.[144]

(121) In the seventh place, the provider of X argues that Regulation (EU) 2022/2065 does not require providers of very large online platforms to ensure the authenticity of users and that X's verification policy is a voluntary commitment to a '*trustworthy*' online environment. According to the provider of X, the aim of the policy change in November 2022 was to reduce fake and untrustworthy accounts.[145]

(122) In the eighth place, the provider of X claims that the Commission aims to impose a specific approach to verification and that there is no legal basis for this in Regulation (EU) 2022/2065. The provider of X argues that this would '*constitute an unwarranted breach of X's freedom to conduct its business and of general principles of freedom of expression of both X and its users.*'[146]

(123) In the ninth place, the provider of X argues that the Commission's reliance on statements from X's Risk Assessment Reports to substantiate findings of non-compliance with Regulation (EU) 2022/2065 sets adverse incentives for providers of very large online platforms and of very large online search engines to include less details in their risk assessment reports.[147]

(124) In the tenth place, the provider of X claims that the Commission '*does not seem to have taken into consideration the exculpatory information available in the case file.*'[148] The provider of X questions the objectiveness, thoroughness, and impartiality of the Commission's investigation.[149]

4.5. The Commission's assessment of the provider of X's arguments

(125) For the reasons set out in the following recitals, the Commission finds that the arguments put forward by the provider of X do not call into question the Commission's finding of an infringement of Article 25(1) of Regulation (EU) 2022/2065 by that provider, as set out in section 4.3 above.

4.5.1. The claims regarding a lack of legal clarity and guidance

(126) In the first place, the Commission considers that the provider of X's arguments regarding a lack of legal clarity and guidance in relation to Article 25(1) of Regulation (EU) 2022/2065 are unfounded.

---

143 Reply to the Preliminary Findings, paragraphs 92-93.
144 Reply to the Preliminary Findings, paragraphs 152-157, 161-163.
145 Reply to the Preliminary Findings, paragraphs 90, 95-96.
146 Reply to the Preliminary Findings, paragraphs 104-106, 165-173.
147 Reply to the Preliminary Findings, paragraph 159.
148 Reply to the Preliminary Findings, paragraphs 88, 111-112.
149 Reply to the Preliminary Findings, paragraphs 89, 112, 139-140, 151, 164.

(127) First, as regards the provider of X's claim that the Commission should have issued guidelines on Article 25(1) of Regulation (EU) 2022/2065 before launching an investigation into its non-compliance with that provision, the Commission observes that Article 25(3) of that Regulation merely empowers, but does not oblige, the Commission to issue guidelines on the practical application of Article 25(1).

(128) Pursuant to Article 288 of the Treaty on the Functioning of the European Union, Regulation (EU) 2022/2065 has '*general application*' and is '*binding in its entirety and directly applicable in all Member States*'. By its very wording, Article 25(1) of Regulation (EU) 2022/2065, which is phrased as a prohibition, is self-executing and thus directly applicable. Article 25(1) of Regulation (EU) 2022/2065 as clarified by the related recitals, clearly set out the scope of the provider's obligations under that provision, as well as the intention of the Union legislature in adopting that provision. Consequently, the Commission's power to adopt guidelines on Article 25(1) of Regulation (EU) 2022/2065, as provided by Article 25(3) of that Regulation, does not preclude the direct applicability of that provision in the absence of such guidance, nor could it, since guidelines are non-binding instruments.

(129) The Commission considers that, were it required to issue guidance before taking any enforcement steps (*quod non*), it would be effectively prevented from enforcing obligations laid down by Regulation (EU) 2022/2065 for a substantial period of time, leading to a delay in the full achievement of the objectives that that Regulation aims to achieve. This would give rise to a risk of potentially allowing an online environment which threatens the fundamental rights provided for in the Charter to persist and develop.[150] In this regard, the Commission notes that the Union legislature attached particular importance to applying Regulation (EU) 2022/2065 as quickly as possible to providers of very large online platforms and of very large online search engines, as demonstrated by the advanced entry into force of that Regulation for those providers.[151]

(130) Second, as regards the provider of X's claim that the Commission's interpretation of Article 25(1) of Regulation (EU) 2022/2065 stretches of the notion of 'decisions' used in that provision, the Commission rejects that claim for following reasons.

(131) First of all, as stated in Article 25(2) of Regulation (EU) 2022/2065, the prohibition in paragraph 1 of that provision does not apply to practices covered by Directive 2005/29/EC. It therefore follows that the notion of 'decision' used in Article 25(1) of Regulation (EU) 2022/2065 necessarily covers a broader set of practices than the transactional economic behaviour covered by Directive 2005/29/EC, which is presumably what the provider of X means when it refers to the '*natural meaning and intended confines'* of the notion of 'decisions'.[152] As such, Article 25(1) of Regulation (EU) 2022/2065, unlike Directive 2005/29/EC, '*applies to all recipients of online platforms, irrespective of whether they are consumers, and safeguards their ability to make free and informed decisions in relation to all aspects of online platforms.*'[153]

---

150 Order of the Vice President of the Court 24 March 2024, Amazon Services Europe v Commission, C-639/23, ECLI:EU:C:2024:277, paragraph 157.

151 Order of the Vice President of the Court, Amazon Services Europe v Commission, C-639/23, ECLI:EU:C:2024:277, paragraph 162.

152 Reply to the Preliminary Findings, paragraph 144 (DSA.100101, Doc ID 258-2).

153 Wilman, Folkert, et al. The EU Digital Services Act, Oxford University Press, Incorporated, 2024. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/europaeu/detail.action?docID=31531943, page 201.

This includes, but is not limited to, decisions of recipients on the authenticity of the accounts whose content they consume on a social media service advertising itself as a source for information and news, such as the service by the provider of X.

(132) In addition, while the Commission agrees that Article 25(1) of Regulation (EU) 2022/2065 should not be read as a general regulatory burden for any design element of minor relevance to the core functioning of a service and to how recipients use that service, in the present case, the deceptive character of the 'verified' status is of central relevance to what the provider of X claims is a core function of the X service as a source for news and information, where source verification is an integral part of the user experience and X's value proposition. Without prejudice to the crucial role that online anonymity plays in enabling dissenting voices to speak freely, *deceptive* claims as to the 'verified' status of accounts have the potential to deceive users about the authenticity of those accounts and the content they encounter on those accounts or in relation to them, which in turn distorts or impairs their ability to make free and informed decisions regarding the authenticity and reliability of those accounts.

4.5.2. The claims regarding alleged general awareness of new meaning of 'verified' status

(133) In the second place, the Commission considers that the provider of X's argument regarding the alleged awareness of the average X user regarding the changed meaning of the 'verified' status is unfounded.

(134) First, that argument rests on a false equivalence between a time-limited awareness in certain sections of the media and the general awareness of X's more than 100 million users in the Union. Neither the specific media articles, nor X's own public announcements to which the provider of X refers, can be assumed to have reached and to have been fully understood by the entirety or even a substantial part of the recipients of the X service.

(135) The Commission notes, in this regard, that those media articles and announcements are mostly in English, whereas the X service is potentially used in all 24 official languages of the Union. Moreover, both the announcements by the provider of X and the media attention given to the issue were largely limited to the time-period from 2023 to early 2024, whereas the deceptive design practices regarding the misleading 'verified' status persist to this day.

(136) Second, it is conceptually flawed to argue that time-limited media attention and linguistically limited one-off announcements by the provider of X on the 'verified' status could undo or mitigate the deceptive nature of the new 'verified' status in the light of Article 25(1) of Regulation (EU) 2022/2065. Were the Commission to accept such an argument, it would mean that *prima facie* deceptive practices that receive sufficient media attention would escape the prohibition in Article 25(1) of Regulation (EU) 2022/2065 and that limited corrective statements by the provider of an online platform could undo the *prima facie* deceptive nature of relevant elements of that platform's online interface design. This argument is particularly flawed, since media attention and the need for corrective statements might be a direct consequence of the blatantly deceptive character of the feature of the platform.

(137) The media articles to which the provider of X refers and the announcements made by that provider on the new 'verified' status in this respect are therefore largely irrelevant to assess that provider's compliance with the prohibition laid down in Article 25(1) of Regulation (EU) 2022/2065. New recipients of the X service, or those recipients that did not pay attention to either media coverage or announcements made by the provider

of X at the time, are left with nothing more than the provider of X's deficient disclaimers about the deceptive claim and design of the current 'verified' status.

(138) As regards the provider of X's help pages containing explanatory information on the new meaning of the 'verified' status, as described in recitals 103 to 105 above, the Commission has acknowledged the existence of these pages and found them to be inadequate to undo the *prima facie* deceptive nature of how that status is generally presented on X's online interface. The Commission has demonstrated that those pages are significantly removed from the actual user experience of the deceptive practice of the 'verified' status, so that the provider of X's extensive cataloguing and quoting of its various help pages is unable to call into question the Commission's preliminary findings.

4.5.3. The claims that the Commission's evidence is insufficient

(139) In the third place, the Commission considers that the provider of X's argument regarding the alleged lack of evidence establishing non-compliance with Article 25(1) of Regulation (EU) 2022/2065 is unfounded.

(140) Requiring the Commission to conduct statical consumer testing and/or to send requests for information to third parties on the risks of deception of the average EU user to establish an infringement of Article 25(1) of Regulation (EU) 2022/2065 in all cases would impose a disproportionate burden of proof in cases of straightforward infringements, as in the present case. As laid out in the Commission's assessment in section 4.3. above, it is obvious that conferring the 'verified' status to accounts that have, in fact, not been verified is a deceptive practice which materially distorts or impairs the ability of recipients of X's service to make decisions in relation to the authenticity of the content they are exposed to on that service. The publicly available and open-source information, referred to in section 4.3, is in itself sufficient to support the Commission's assessment of the existence of an infringement of Article 25(1) of Regulation (EU) 2022/2065 such as the one in the present case.

(141) Furthermore, the provider of X's claims are not supported by existing legal frameworks. As regards the Commission's assessment of the legal framework as described by the provider of X, recital 18 of Directive 2005/29/EC[154] expressly states that *'[t]he average consumer test is not a statistical test. National courts and authorities will have to exercise their own faculty of judgement, having regard to the case-law of the Court of Justice, to determine the typical reaction of the average consumer in a given case*'.[155] Furthermore, as stated in the Guidance on Unfair Commercial Practices *'[t]his means that national authorities and courts should be able to determine whether a practice is liable to mislead the average consumer exercising their own judgment by taking into account the general presumed consumers' expectations, without having to commission an expert's report or a consumer research poll.*'[156] In analogy to recital 18 of Directive 2005/29/EC, as read

154 Recital 18 of Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65 of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('The Unfair Commercial Practices Directive'), OJ L 149, 11.6.2005.

155 Recital 18 of Directive 2005/29/EC Unfair Commercial Practices Directive.

156 Commission Notice on Guidance on the interpretation and application of Directive 2005/29/EC of the European Parliament and of the Council concerning unfair business-to-consumer commercial practices

in light of the aforementioned Guidance, the Commission maintains that the necessary scope and evidentiary benchmark for establishing a deceptive practice under Article 25(1) of Regulation (EU) 2022/2065 is dependent on the details of the specific case and proportionate to the infringement.

4.5.4. The claims questioning the probative value of the survey study

(142) In the fourth place, the Commission considers the provider of X's argument regarding the alleged lack of probative value from the survey study on 'Account Verification on Social Media' invoked in the Preliminary Findings is unfounded.

(143) First, some of the provider of X's claims are misguided insofar as they suggest that the probative value of evidence should be tied to academic seniority or that only evidence that has gone through the full academic peer review cycle is relevant for the assessment of whether a certain practice infringes Regulation (EU) 2022/2065. Rather, evidence should be weighed on the basis of its methodological soundness and factual relevance to the interpretation of Regulation (EU) 2022/2065.

(144) Second, while the Commission referenced the pre-print version of the aforementioned study in the Preliminary Findings, that study has since been accepted as a conference paper at the 'SEC '23: Proceedings of the 32nd USENIX Conference on Security Symposium' of 9 August 2023, which included a selective peer-review process, a public fact that the provider of X could have been aware of at the time of replying to the Preliminary Findings.[157]

(145) In any event, while any sample-based study on any issue has inferential limitations, the aforementioned study is of probative value insofar as it diligently summarises historical and cross-industry practices and provides a relevant statistical consumer test of relevance for the Commission's assessment of the provider of X's compliance with Article 25(1) of Regulation (EU) 2022/2065. Moreover, contrary to the suggestion by the provider of X, the study is only one of many pieces of evidence relied upon by the Commission to assess the provider of X's compliance with Article 25(1).

(146) As regards the provider of X's claim that the Commission wrongfully ignored a reference to another study from the bibliography of the aforementioned study,[158] it is irrelevant, since the referenced study is about whether users perceive content from verified accounts to be more credible. The Commission's assessment of X's 'verified' status feature under Article 25(1) of Regulation (EU) 2022/2065 does not deal with the credibility – or veracity – of the content posted by 'verified' accounts, but with the deceptiveness of the feature itself. Insofar as the present case is about the legitimacy of *authenticity* claims regarding the identity of a user account, the referenced study is irrelevant.

---

in the internal market ("*Notice on the Unfair Commercial Practices Directive"*), OJ C 526, 29.12.2021, page 33.

157 Xiao, Madelyne, Mona Wang, Anunay Kulshrestha, and Jonathan Mayer. 2023. Account verification on social media: user perceptions and paid enrollment. In Proceedings of the 32nd USENIX Conference on Security Symposium (SEC '23). USENIX Association, USA, Article 174, 3099–3116. DSA.100101, Doc-ID 122-1.

158 Vaidya, Tavish, Daniel Votipka, Michelle L. Mazurek, and Micah Sherr. 2019. Does Being Verified Make You More Credible? Account Verification's Effect on Tweet Credibility. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). Association for Computing Machinery, New York, NY, USA, Paper 525, 1–13. https://doi.org/10.1145/3290605.3300755 DSA.100101, Doc ID 161.

4.5.5. The claims regarding alleged shortcomings of Twitter's legacy verification policy

(147) In the fifth place, as regards the imperfections of the previous verification program, the provider of X's arguments are unfounded and the facts about the legacy program the provider of X raised are ultimately inculpatory for the present case.

(148) The Commission never found that the provider of Twitter's legacy verification program was failure-proof, but rather that the provider of Twitter engaged in a good faith effort to reliably verify accounts to which it afforded the verified status. This stands in stark contrast to the minimal and ultimately ineffective requirements implemented by the provider of X's new approach to the 'verified' status. That contrast is elaborated in detail in the Commission's assessment in section 4.3 above.

(149) Furthermore, the fact that the provider of X mentions how Twitter rescinded its wider roll-out of the verification program in May 2021, due to scaling issues for providing a reliable vetting process that upholds the credibility of the 'verified' status, is further proof that the provider of X is aware of the serious risks and challenges related to a generalised public verification program. In the light of this awareness, it is even more concerning that, as laid out in recitals 80 to 83 above, the provider of X has not developed any meaningful processes or dedicated meaningful resources to reliably verify applications for its current 'verified' status program.

4.5.6. The claims regarding alleged inevitability of verification mistakes

(150) In the sixth place, the arguments of the provider of X as regards inevitable imperfections in any verification process, mitigation measures put in place and the prohibition of a general monitoring obligation are ineffective.

(151) In finding that the provider of X has infringed Article 25(1) of Regulation (EU) 2022/2065, the Commission is not imposing a general monitoring obligation on that provider. Rather, the Commission has assessed whether the 'verified' status feature is designed, organised or operated in such a way that deceives the recipients of X thereby materially distorting or impairing the ability of those recipients to make free and informed decisions regarding the authenticity and reliability of those accounts. The Commission's finding that it does, necessarily means that the Commission does not consider the so-called 'mitigation measures' the provider of X has put into place to remedy the deceptive design of the feature.

4.5.7. The claims regarding alleged benefits of X's approach to the 'verified' status

(152) In the seventh place, the provider of X's arguments regarding the alleged benefits of its 'verified' status are irrelevant to the present case.

(153) The Commission's preliminary findings regarding X's non-compliance with Article 25(1) of Regulation (EU) 2022/2065 do not stem from whether X has a verification policy in place, but whether the policy it has in place is designed in such a manner so as to deceive the recipients of X thereby materially distorting or impairing the ability of those recipients to make free and informed decisions regarding the authenticity and reliability of those accounts.

(154) Even if the provider of X's claim that the purpose of its policy was to reduce fake and untrustworthy accounts were true (*quod non*), that would only reduce the statistical likelihood of inauthentic accounts having a 'verified' status, without eliminating that possibility. Such a statistical reduction might be a reasonable and even desirable approach, but it is irrelevant to the Commission's assessment of the provider of X's compliance with Article 25(1) of Regulation (EU) 2022/2065. That assessment takes

issue with the policy's absolute and *ex ante* claim to authenticity of the 'verified' status, which is deceptive to the recipients of X insofar as fake and untrustworthy accounts are still able to receive the 'verified' status, as explained in the Commission's assessment in section 4.3 above.

(155) As such, the assessment of this policy under Article 25(1) of Regulation (EU) 2022/2065 in no way precludes the Commission's assessment of the 'verified' status in the context of the obligations to perform risk assessments and to put in place mitigation measures in relation to systemic risks to civic discourse and electoral processes pursuant to Article 34(1) and (2) and Article 35(1) of Regulation (EU) 2022/2065, which are the subject-matter of case DSA.100100.

4.5.8. The claims regarding alleged imposition of a specific approach to verification

(156) In the eighth place, the provider of X's claim that the Commission is trying to impose a specific verification approach is unfounded.

(157) The Commission's preliminary finding in the present case is that the provider of X infringed Article 25(1) of Regulation (EU) 2022/2065 by combining a lenient vetting system that is significantly less rigorous than that previously in place with the same historical online interface design that used to indicate that the identity and authenticity of an account holder was 'verified'. Such a finding does not impose an obligation to verify users, nor any specific approach to verification. In any event, the provider of X's freedom to conduct a business is not absolute. That freedom must, first, be viewed in relation to its function in society and, second, be weighed in the balance with other interests protected by the EU legal order and the rights and freedoms of others.[159] In any event, that freedom may be subject to prohibitions laid down by law which are necessary and proportionate, as is the case for the prohibition laid down in Article 25(1) of Regulation (EU) 2022/2065.

4.5.9. The claims regarding alleged adverse incentives of reliance on Risk Assessment Reports

(158) In the ninth place, the provider of X's arguments regarding potential adverse incentives stemming from the Commission's reliance on evidence from its risk assessment report are unfounded and irrelevant for the present case.

(159) At the outset, the Commission observes that it has not simply relied on the summary report of X's Risk Assessment Report, but has followed it up with several RFIs, asking both for supporting documentation pursuant to Article 34(3) of Regulation (EU) 2022/2065 and for replies in relation to specific concerns regarding compliance by the provider of X with Article 25(1) of that Regulation in relation to the 'verified' status. The Commission has therefore given the provider of X ample opportunity to dispel the Commission's concerns and/or to address shortcomings identified in its Risk Assessment Report.

(160) Next, the risk assessment performed pursuant to Article 34 of Regulation (EU) 2022/2065 provides an opportunity for providers of very large online platforms and of very large online search engines to document and communicate their compliance efforts, which is a crucial aspect for the effectiveness of Regulation (EU) 2022/2065. The provider of X has admitted in its Risk Assessment Report that it is aware of

[159] Case C-124/20, Bank Melli Iran, EU:C:2021:1035, paragraphs 80-83.

certain risks related to the 'verified' status, but it has failed to respond in a way that suggests that it is serious about addressing those concerns.

(161) Finally, prohibiting the Commission from taking into account a provider's risk assessment and a provider's explanations for the purposes of supervising and enforcing Regulation (EU) 2022/2065 would set perverse incentives for providers of very large online platforms and of very large online search engines to regard their obligations under Article 34 of Regulation (EU) 2022/2065 as a mere formality.

4.5.10. The claims regarding alleged exculpatory information and partiality of investigation

(162) In the tenth place, the provider of X's arguments on the alleged partiality of the Commission's assessment and the alleged omission of exculpatory information on file are unfounded and misleading.

(163) First of all, the alleged 'exculpatory information' is irrelevant to the case at hand. At a very early stage of the investigation, the Bundesnetzagentur ('BNetzA'), i.e. the German Digital Service Coordinator ('DSC'), informed the Commission that, in its cursory assessment of evidence specific for the situation in Germany, it had found a few cases of satirical accounts of German politicians that had the 'verified' status, but were clearly marked as satire. Because of the cursory and preliminary nature of the assessment, the Commission did not consider it relevant for the Preliminary Findings, also because the Commission's finding of an infringement of Article 25(1) of Regulation (EU) 2022/2065 is neither about satirical accounts nor impersonation issues as such.

(164) Furthermore, the BNetzA claimed that '*[i]n favour of IUC and X Holdings Corp. it has to be pointed out that the explanation on what the blue checkmark is and the requirements for the assignment are transparent. Information regarding this is freely accessible on the help centre pages. Additionally, the media has frequently discussed the subject, so that a certain awareness should be present*.' However, it also clearly stated that '*while BNetzA endeavours to support the proceedings by sharing relevant external information, it encourages the EU Commission to independently verify the accuracy of the provided details.*'[160]

(165) With appreciation for BNetzA's expeditious input to the investigation, the Commission indeed independently verified its initial suspicions and found in the Preliminary Finding that, considering the additional evidence gather by the Commission in the course of its in-depth investigation, as set out and discussed in the Preliminary Findings, the fact that some information on X's help pages is available is in itself insufficient to render the online interface design related to the 'verified' status non-deceptive. The Commission did not deem the mere mentioning of the help page by this early assessment of the BNetzA relevant enough to be considered 'exculpatory information' that ought to be included in the assessment, which already extensively discusses the presence and use of the help page information.

4.6. Conclusion

(166) For all the reasons set out in sections 4.3 and 4.5, the Commission concludes that the provider of X has failed to comply with Article 25(1) of Regulation (EU) 2022/2065 by designing, organising and operating X's online interface in relation to the 'verified' status feature of that service in a manner that deceives recipients of its service into

[160] DSA.100101, Doc ID 211.

believing that accounts with a 'verified' status on X are meaningfully vetted by that provider thereby materially distorting or impairing their ability to make free and informed decisions regarding the authenticity and reliability of those accounts. . This conclusion does not prejudge the outcome of the Commission's ongoing investigation into the compliance by the provider of X with its obligations stemming from Article 34(1) and (2), and Article 35(1) of Regulation (EU) 2022/2065 in case DSA.100100 – X (formerly Twitter), in relation to the risk assessment and the effectiveness of mitigation measures linked to subscription services, such as the X Premium and Premium+ subscription services.

5. NON-COMPLIANCE WITH ARTICLE 39 OF REGULATION (EU) 2022/2065

5.1. The legal framework

(167) Pursuant to Article 39(1) of Regulation (EU) 2022/2065, '[*p*]*roviders of very large online platforms and of very large online search engines that present advertisements on their online interfaces shall compile and make publicly available in a specific section of their online interface, through a searchable and reliable tool that allows multicriteria queries and through application programming interfaces, a repository containing the information referred to in paragraph 2, for the entire period during which they present an advertisement and until one year after the advertisement was presented for the last time on their online interfaces. They shall ensure that the repository does not contain any personal data of the recipients of the service to whom the advertisement was or could have been presented, and shall make reasonable efforts to ensure that the information is accurate and complete*.'

(168) Pursuant to Article 39(2) of Regulation (EU) 2022/2065, providers of very large online platforms and of very large online search engines are required to ensure that the advertising repository shall contain the following information to facilitate research and supervision by researchers and regulators: *'(a) the content of the advertisement, including the name of the product, service or brand and the subject matter of the advertisement; (b) the natural or legal person on whose behalf the advertisement is presented; (c) the natural or legal person who paid for the advertisement, if that person is different from the person referred to in point (b); (d) the period during which the advertisement was presented; (e) whether the advertisement was intended to be presented specifically to one or more particular groups of recipients of the service and if so, the main parameters used for that purpose including where applicable the main parameters used to exclude one or more of such particular groups; (f) the commercial communications published on the very large online platforms and identified pursuant to Article 26(2); (g) the total number of recipients of the service reached and, where applicable, aggregate numbers broken down by Member State for the group or groups of recipients that the advertisement specifically targeted.*'

(169) Recital 68 of Regulation (EU) 2022/2065 stresses the importance of transparency concerning online advertisements in the online environment noting that '[*o*]*nline advertising can contribute to significant risks, ranging from advertisements that are themselves illegal content, to contributing to incentives for publication or amplification of illegal or otherwise harmful content and activities online, or the discriminatory presentation of advertisements with an impact on the equal treatment and opportunities of citizens'*.

(170) Article 39 of Regulation (EU) 2022/2065 is one of several transparency obligations for providers of very large online platforms and of very large online search engines laid

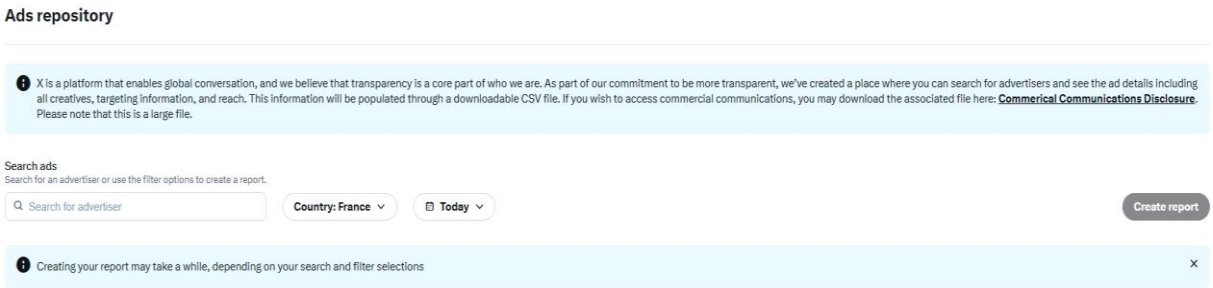
down in that Regulation, including transparency reporting,[161] the publication of risk assessment and risk mitigation measures reports[162] and independent audit report,[163] and data access for researchers.[164] Recital 40 of that Regulation emphasises that these obligations intend to ensure the safety and trust of the recipients of the service, the meaningful accountability of those providers, and the empowerment of recipients and other affected parties, while facilitating the necessary oversight by competent authorities.

(171) Recital 95 of that Regulation adds that advertising systems used by providers of very large online platforms and of very large online search engines pose particular risks, requiring '*further public and regulatory supervision on account of their scale and ability to target and reach recipients of the service based on their behaviour within and outside that platform's or search engine's online interface'*. For that reason, providers of very large online platforms and of very large online search engines '*should ensure public access to repositories of advertisements presented on their online interfaces to facilitate supervision and research*'. This will facilitate the dual objective of facilitating supervision and research '*into emerging risks brought about by the distribution of advertising online, for example in relation to illegal advertisements or manipulative techniques and disinformation with a real and foreseeable negative impact on public health, public security, civic discourse, political participation and equality. Repositories should include the content of the advertisements, including the name of the product, service or brand and the subject matter of the advertisement, and related data on the advertiser, and, if different, the natural and legal person who paid for the advertisement, and the delivery of the advertisement, in particular where targeted advertisement is concerned. This information should include both information about targeting criteria and delivery criteria in particular when advertisements are delivered to persons in vulnerable situations, such as minors.*'

5.2. The relevant measures put in place by the provider of X

(172) To fulfil its obligations under Article 39 of Regulation (EU) 2022/2065, the provider of X made publicly available on a specific section of X's online interface a repository of advertisements displayed on X.[165] Figure 4 below provides a visual representation of that advertising repository.

Figure 4: The provider of X's advertising repository 14 February 2025.



Source: X website, https://ads.x.com/ads-repository, accessed on 19 June 2024, (DSA.100103 160-29).

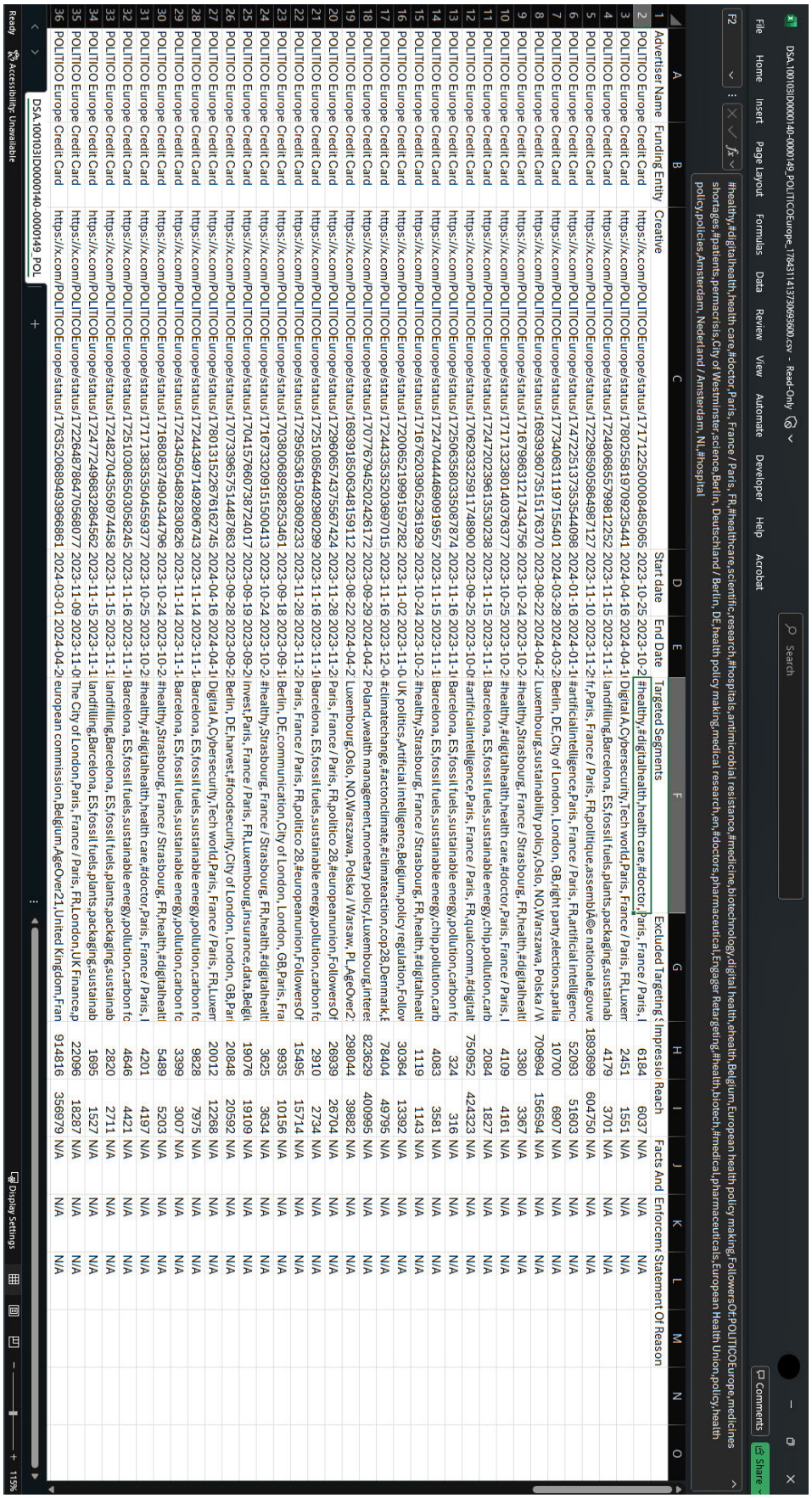
161 Article 42, Regulation (EU) 2022/2065.
162 Article 42(4)(b), Regulation (EU) 2022/2065.
163 Article 37, Regulation (EU) 2022/2065.
164 Article 40, Regulation (EU) 2022/2065.
165 'DSA.100103' Technical Analysis, section 1.1 (DSA.100103, Doc ID 219-1).

(173) X's advertisement repository is accessible through a search interface that enables users to search for advertisements based on three criteria: (i) the advertiser's X handle; (ii) the individual Member State in which the advertisement was displayed; and (iii) the time frame during which the advertisement was active. Figure 4 above provides a visual representation of those search criteria.

(174) To generate an advertising report from the repository, all fields must be defined by the user.[166] That report is then generated and presented as a comma--separated values ('.csv') file which requires third-party software (such as Excel) to be viewed.[167] The .csv file generates the following fields: (i) the advertiser's name; (ii) the entity funding the advertisement; (iii) a link to the advertisement content; (iv) the start date of the advertisement campaign; (v) the end date of that campaign; (vi) the segments targeted; (vii) the segments excluded from targeting; (viii) advertisement impressions; and (ix) advertising reach.[168] Figure 5 below provides a visual representation of .csv file results.

166 'DSA.100103' Technical Analysis, section 1.1 (DSA.100103, Doc ID 219-1).
167 'DSA.100103' Technical Analysis, section 1.1 (DSA.100103, Doc ID 219-1).
168 Reply to the Preliminary Findings, paragraph 184.

Figure 5: .csv file taken 19 June 2024.

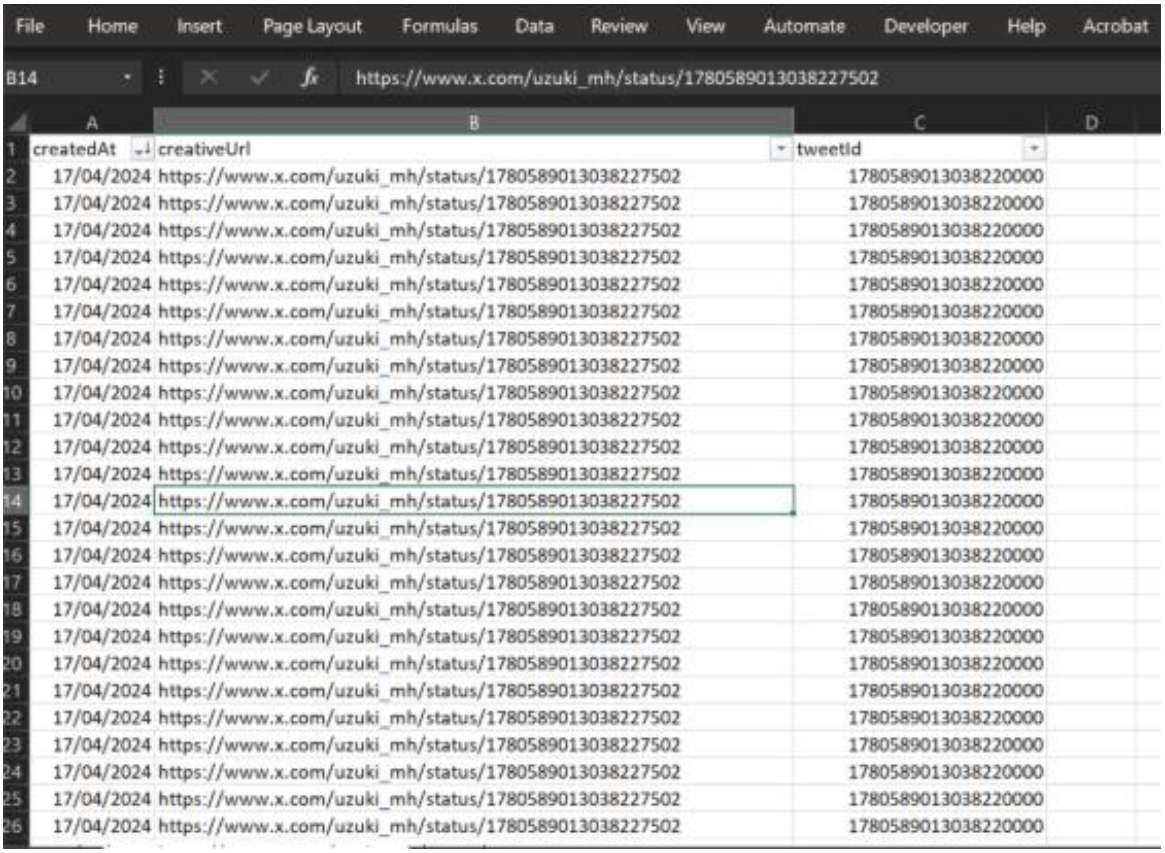


Source: Csv from the provider of X's advertising repository https://ads.x.com/ads-repository (source: ECAT, last accessed 19 June 2024, DSA.100103, Doc ID 219).

2959

(175) Information concerning commercial communications is not directly searchable through the main search tool made available on X's online interface. Instead, users must download a separate .csv file.[169] Figure 6 below provides a visual representation of the Commercial Communications .csv file.

Figure 6: Commercial communications .csv file taken 10 June 2024.

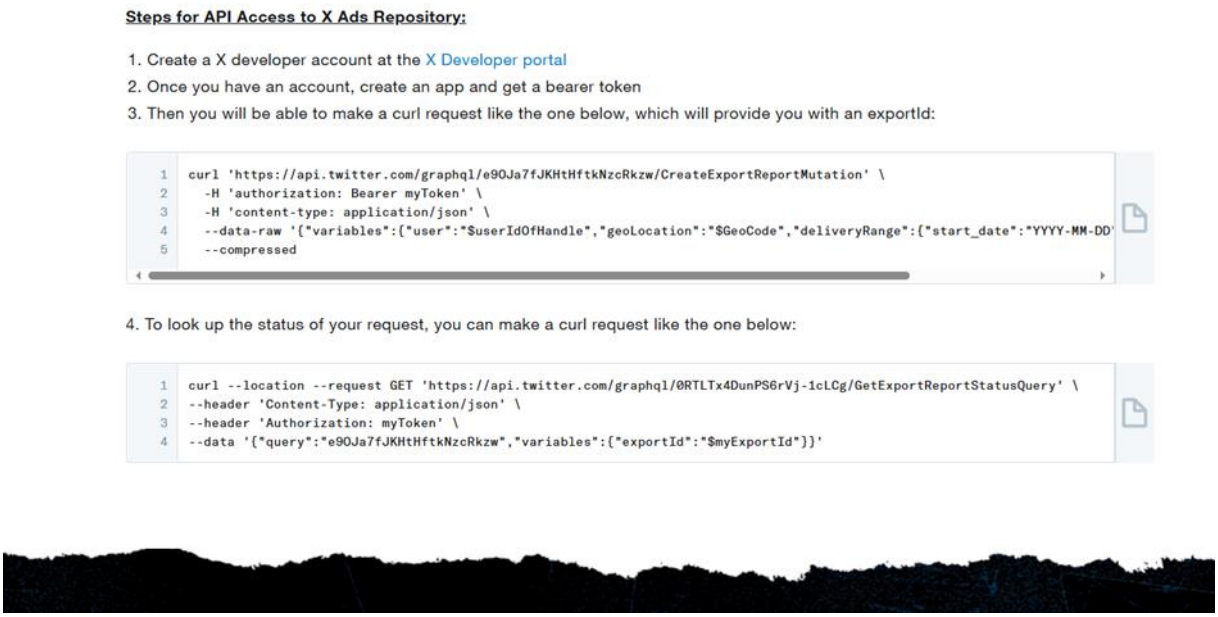


Source: X website, https://ads.x.com/ads-repository, accessed on 19 June 2024, (DSA.100103, Doc ID 219).

(176) Instructions on how to gain access to the information in the repository through an application programming interface ('API') are provided in a separate section of X's online interface.[170] Figure 7 below provides a visual representation of API access steps.

169 'DSA.100103' Technical Analysis, section 1.1 (DSA.100103, Doc ID 219-1).
170 'DSA.100103' Technical Analysis, section 1.2 (DSA.100103, Doc ID 219-1).

Figure 7: Steps for API access taken 14 February 2025.



Source: X website, https://business.x.com/en/help/ads-policies/product-policies/ads-transparency accessed 19 June 2024 (DSA.100103 Doc ID 219).

5.3. The Commission's Preliminary Findings

(177) In its Preliminary Findings, the Commission reached the preliminary conclusion that the provider of X failed to comply with Article 39 of Regulation (EU) 2022/2065 by failing to put in place an advertisement repository that meets the requirements set out in that provision. In particular, the advertisement repository put in place by the provider of X does not provide a searchable and reliable tool that allows multicriteria queries, nor an API enabling public access to all the information referred to in Article 39(2) of Regulation (EU) 2022/2065, in relation to advertisements that the provider presents on X's online interface. That conclusion was based on the following considerations.

5.3.1. X's advertising repository is not available through a searchable tool

(178) In the first place, the Commission considered that the tool[171] through which X has made its advertisement repository available ('the search tool') does not constitute a '*searchable […] tool that allows multicriteria queries*' within the meaning of Article 39(1) of Regulation (EU) 2022/2065.

(179) First, the search tool does not allow multicriteria queries in a way that would fulfil the objectives of that provision to facilitate supervision and research into emerging risks brought about by the distribution of advertising online. Trial interactions with X's advertisement repository conducted by the Commission[172] show that queries for downloading data on advertisements presented on X are limited to the following search fields: (i) the Member States where the advertisement was presented (a user must select one Member State, with no possibility to choose the whole of the Union); (ii) the X account of an advertiser (a user must select one advertiser from a drop-down list displayed); and (iii) the time frame within which the advertisement was presented.

171 X website, https://ads.x.com/ads-repository, accessed on 19 June 2024 (DSA.100103, Doc ID 160-29).
172 'DSA.100103' Technical Analysis, section 1.1 (DSA.100103, Doc ID 219-1).

This is also clear from Figure 7 above, which reproduces the interface to X's advertisement repository. To perform a search and obtain an advertisement report, all of the aforementioned search fields must be filled in, which means that it is not possible to search by selecting only one of those fields.

(180) To ensure compliance with Article 39(1) of Regulation (EU) 2022/2065, the search tool should allow multicriteria search queries covering at least all the information listed in Article 39(2) of that Regulation. Only in this manner can the repository serve its intended purpose of facilitating supervision and research into emerging risks brought about by the distribution of advertising online.

(181) Due to the way in which the search tool has been implemented by the provider of X, a user would not be able to directly perform multicriteria searches using X's repository based on:

(1) the content of the advertisement, including the name of the product, service or brand and the subject matter of the advertisement;

(2) the natural or legal person who paid for the advertisement;

(3) whether the advertisement was intended to be presented specifically to one or more particular groups of recipients of the service and, if so, the main parameters used for that purpose including where applicable the main parameters used to exclude one or more of such particular groups; or

(4) the commercial communications published on the service.

(182) At present, most of the information listed in Article 39(2) of Regulation (EU) 2022/2065 can only be obtained indirectly from X's advertisement repository by searching a single advertiser in a single Member State at a particular moment in time. This does not facilitate the supervision and research into emerging risks stemming from online advertisements concerning risks related to public security, civic discourse, political participation or equality, as required by that Article.

(183) Second, the way in which the results of the search query are displayed does not allow public scrutiny in line with the objectives of Article 39 of Regulation (EU) 2022/2065 to facilitate supervision and research into emerging risks brought about by the distribution of advertising online.

(184) More specifically, those search results are not presented in a specific section of X's online interface, as required by Article 39(1) of Regulation (EU) 2022/2065, but are only downloadable in a separate .csv file, whose search functionalities depend on the use of third-party software to open the file. The absence of a visual display of the query results in X's online interface with search and sorting functions further reduces the searchability of the tool, since search results can only be accessed with third-party software that opens the .csv file. This applies to information on commercial communications as well, which are excluded altogether from the tool's search functions and are only accessible through a separate .csv file.[173]

(185) These limitations to the search functionalities and the inadequate manner in which the results can be retrieved frustrate the purpose of Article 39 of Regulation (EU) 2022/2065 to facilitate supervision and research into emerging risks of online

[173] 'DSA.100103' Technical Analysis, section 1.1 (DSA.100103, Doc ID 219-1).

advertising. This is further corroborated by experts consulted by the Commission during its investigation into X's advertisement repository.

(186) For example, a representative from CheckFirst, a research company that together with the Mozilla Foundation published a report on [redacted] 2024[174] concerning advertisement repositories, pointed out that the requirement of a searchable and reliable tool set out in Article 39(1) of Regulation (EU) 2022/2065 should enable multicriteria searches to find advertisements that would otherwise remain undiscovered on the platform. In contrast, X's advertisement repository requires a user to have already identified the advertisement that they are looking for in order to perform a search query.[175] That representative further pointed out that providing results to a search query in a .csv file, instead of on the platform's user interface, limits the accessibility and ability to visualise those results. Similarly, a representative of Ravineo, a company that works on advertisement transparency, highlighted that the current absence of effective multi-criteria search query capacities in X's advertisement repository is a limiting factor in finding the information necessary to compare political advertisement campaigns, and contrasted X's advertisement repository with better alternatives that enable searches by advertising accounts or full-text queries.[176]

(187) An illustrative example of deceptive advertising that has been detected on Facebook and Instagram, but would not be detectable from X's advertisement repository due to its limited search functionality, is provided in a study by AI Forensics analysing disinformation campaigns, likely of Russian origin, on Meta ahead of EU parliamentary elections in 2024.[177] According to the aforementioned expert from CheckFirst, carrying out equivalent research on X would require to extend the searchability functions of X's advertisement repository to ensure that advertisement campaigns currently under the radar are discovered, possibly across several Member States.[178]

(188) Third, trials conducted by the Commission[179] demonstrate that X's search tool produces the .csv advertisement reports to each search query only after a response time of three minutes and twenty seconds. This very slow response time reduces the effectiveness of the search tool for its intended purpose to facilitate research and supervision into emerging risks from online advertising, making it de facto unusable in practice.

(189) The aforementioned trials provided no apparent technical reason or justification for this slow response time. The trials demonstrated that – with significant technical knowhow applied – each report could actually be obtained within a few seconds, showing that the slow response time is artificially increased.[180] The Commission

---

174 Mozilla Foundation and CheckFirst, Full Disclosure: Stress testing tech platforms' ad repositories, 16 April 2024 (DSA.100103, Doc ID 171-1).

175 Non-confidential minutes of interview with a representative from CheckFirst, 15 May 2024 (DSA.100103, Doc ID 166-1).

176 Non-confidential minutes of interview with a representative from Ravineo, 22 April 2024 (DSA.100103, Doc ID 94-1).

177 Bouchaud, Paul, Marc Faddoul, and Raye Buse Çetin. No Embargo in Sight: Meta Lets Pro-Russia Propaganda Ads Flood the EU. 17 April 2024 (DSA.100103, Doc ID 218-1).

178 Non-confidential minutes of interview with a representative from CheckFirst, 15 May 2024 (DSA.100103, Doc ID 166-1).

179 'DSA.100103' Technical Analysis (DSA.100103, Doc ID 219-1).

180 The trials demonstrated that, once a .csv report is requested, the tool waits until the user browser has checked the report status exactly 100 times, whereas it checks for updates with a standard query rate of

considers that the excessively slow response time of the search tool, as compared also with other advertisement repositories and similar databases, stems from its design, and constitutes an unnecessary, intentional, additional barrier to access information on advertisements displayed on X.

(190) This assessment is also confirmed by the external representatives consulted by the Commission. The aforementioned expert from CheckFirst that had analysed and compared all the advertisement repositories of very large online platforms and of very large search engines[181] described X's design of the tool, with the browser side sending regular requests until the reports are produced, as inefficient and pointed out that this design is very unusual as compared to that of other advertisement repositories.[182] This was also noted in Mozilla's and CheckFirst's comparison of various advertisement repositories, as follows: '*our reliability test found that the csv took between five and 10 minutes to load, which is dramatically slower than all of the other platforms studied (most took seconds to respond to a query)*'.[183] The aforementioned representative of Ravineo also stated that, in its view, the constant response time of three minutes and twenty seconds independent of the size of the requested report is not due to technical limitations, but the intentional choice of the provider of X.[184] This view was also reflected by a representative of the Institut des Systèmes Complexes de Paris Ile de France ('ISC-PIF'), who pointed out that the slow response time for reports that are entirely empty seems excessive.[185]

(191) All three experts further expressed the view that the slow response time of X's search tool presents a major obstacle for researchers using the tool for supervision and research into emerging risks brought about by the distribution of advertising online. The representative of CheckFirst stated that the response time makes the scope of any research done on advertisements on X more limited than on other platforms.[186] The representative of Ravineo stated that the slow response time makes it very hard to use the tool to compare the advertisement campaigns of politicians.[187] The representative from ISC-PIF pointed out that long waiting times had impeded their assessment of the accuracy and completeness of X's advertisement repository.[188]

---

two seconds. Only after exactly 100 times, corresponding to three minutes twenty seconds, a report is produced. Those trials further showed that if the query rate is manipulated by the recipient to send the 100 requests in a very short time, the same report can be obtained within a few seconds, demonstrating that the slow response time is artificially increased. This also holds true for requests that concern large online advertisers that ran several hundreds of advertisements on X. 'DSA.100103' Technical Analysis (DSA.100103, Doc ID 219-1).

181 Mozilla Foundation and CheckFirst, Full Disclosure: Stress testing tech platforms' ad repositories, 16 April 2024 (DSA.100103, Doc ID 171-1).

182 Non-confidential minutes of interview with a representative from CheckFirst, 15 May 2024 (DSA.100103, Doc ID 166-1).

183 Mozilla Foundation and CheckFirst, Full Disclosure: Stress testing tech platforms' ad repositories, 16 April 2024, page 41 (DSA.100103, Doc ID 171-1).

184 Non-confidential minutes of interview with a representative of Ravineo, 22 April 2024 (DSA.100103, Doc ID 94-1).

185 Non-confidential minutes of interview with a representative from the Institut des Systèmes Complexes de Paris Ile de France 3 May 2024 (DSA.100103, Doc ID 143-1).

186 Non-confidential minutes of interview with a representative from CheckFirst, 15 May 2024, (DSA.100103, Doc ID 166-1).

187 Non-confidential minutes of interview with a representative of Ravineo, 22 April 2024 (DSA.100103, Doc ID 94-1).

188 Non-confidential minutes of interview with a representative from CheckFirst, 15 May 2024, (DSA.100103, Doc ID 166-1).

(192) In the second RFI, the Commission asked the provider of X to confirm the excessive response time. In its response,[189] the provider of X admitted that:

– '*[…] based on the design of X's ads repository, each report a user requests should take 3.334 minutes to process, which was confirmed by internal testing and is monitored by the relevant engineering team.*

– *Depending on latency and the user's connection, that means the average user should receive their report between three minutes twenty seconds and three minutes forty seconds from the time of their request. This timeframe does not differ based on the number of ads, the country, or the timeframe selected. We do not otherwise maintain an average or median response time for users that would reflect their latency or connection speed for receiving the completed report'*.

(193) This response is contradicted by X's notice to users requesting a report according to which '*creating your report may take a while, depending on your search and filter selections.*'[190] Based on the aforementioned admissions, as well as on the expert testimonies and internal testing and monitoring referred to above, the Commission took the preliminary view that the unnecessarily long response times of the tool are a conscious design decision by the provider of X.

(194) Consequently, the Commission took the view that the provider of X intentionally designed and implemented the search tool of X's advertisement repository in a way that significantly limits the searchability of that repository and its practical usefulness.

5.3.2. X's advertising repository is not available through a reliable tool

(195) In the second place, the Commission considered that the provider of X has not put in place a 'reliable' tool to obtain all the information on advertisements displayed on X required by Article 39(2) of Regulation (EU) 2022/2065.

(196) First, the search tool of X's advertisement repository is not reliable insofar as it does not provide information on all advertisements displayed on X until one year after the advertisement was presented for the last time on X's online interface. Article 39(1) of that Regulation requires provider of very large online platform or online search engine to retain information on the advertisement in their advertisement repositories '*for the entire period during which they present an advertisement and until one year after the advertisement was presented for the last time on their online interfaces.*' (emphasis added)

(197) A report by ISC-PIF[191] demonstrates that, from a random sample of 250 advertisements shown to 151 users of X in France between 1 September 2023 and 18 April 2024, only 58% of those advertisements were found in X's advertisement repository results. According to the aforementioned report, this is largely due to the inability to retrieve information on advertisements from suspended or deleted accounts.

(198) Research carried out by the Commission confirms this reliability issue. In this context, different reports were retrieved from X's advertisement repository on all

189 Reply to the second RFI, section VI, request 1 (b), page 15 (DSA.100103, Doc ID 163-9).

190 'DSA.100103' Technical Analysis, Figure 2 (DSA.100103, Doc ID 219-1).

191 Bouchaud, Paul. X Ads Repository – Evaluating X Ads Repository (In)Completeness Under Article 39 of the Digital Services Act, 8 May 2024 (DSA.100103, Doc ID 100-1).

advertisements by a large advertiser displayed on X in Germany during the relevant period. While a report retrieved on 7 May 2024 still featured over 260 advertisements, a report retrieved on 13 June 2024 concerning the same time period was entirely empty.[192] This was the case, despite the fact that certain advertisements that appeared in the 7 May 2024 report were less than a year old on 13 June 2024, meaning that these advertisements should have additionally appeared in the 13 June 2024 report. These findings were further confirmed by the aforementioned representative from CheckFirst, who acknowledged that there seem to be gaps and inaccurate information if an advertisement has been displayed on X some time ago.[193]

(199) Second, the search tool of X's advertisement repository is not reliable insofar as it does not provide information on the content of the advertisement displayed on X, including the name of the product, service, or brand advertised, and the subject matter of the advertisement.

(200) The reports generated by the search tool include a column entitled '*Creative*', which provides a URL to a post presumably with the content of the advertisement.[194] The reference to such a URL does not amount to the provision of the required information for several reasons.

(201) First of all, the original advertisement may no longer be available under such a URL, as the owner of the post may have deleted or altered it. In addition, due to X's terms of service, a user of the repository may not scrape the URL of a post on X, for example to register the information in the post that allegedly represents the content of the advertisement to complement the downloaded report with this information.[195] Finally, even if a user were able to register the post from a URL, it would be prohibitively difficult to retroactively extract at scale the name of the product, service, or brand advertised and the subject matter of the advertisement, since such information may be contained in text, images, videos or the context of each advertisement. This is the case, in particular, where supervision and research into emerging risks from online advertising rely on the subject of advertisements at scale, for example to exhaustively identify all potentially illegal advertisements. The aforementioned expert from CheckFirst confirmed that the provision of URLs, instead of the actual content of the advertisement, makes it impossible to conduct research relying on the advertisement's content, since this would require an additional research technique or method in combination with a dataset retrieved from the online platform.[196]

(202) Third, the search tool of X's advertisement repository is not reliable insofar as it does not provide information to identify the natural or legal person who paid for the advertisement or, if that person is different, the natural or legal person on whose behalf the advertisement is presented, as required by Article 39(2)(b) of Regulation (EU)

---

192 Repository output for advertisements by HeroWarsWeb in Germany between 28 August 2023 and 13 June 2024 (DSA.100103, Doc ID 206-1).

193 Non-confidential minutes of interview with a representative from CheckFirst, 15 May 2024, (DSA.100103, Doc ID 166-1). This may explains the contrast of the accuracy rate found in the report by the Mozilla and CheckFirst, who checked ten ads shown to one user in the repository the day after they were observed, and the accuracy rate calculated by the expert researcher from ISC-PIF based on a random sample of 250 ads out of 31,362 ads shown to a pool of 151 French users over a period of six months.

194 'DSA.100103' Technical Analysis, Figure 4 (DSA.100103, Doc ID 219-1).

195 X website, https://x.com/en/tos, accessed on 19 June 2024 (DSA.100103, Doc ID 160-65).

196 Non-confidential minutes of interview with a representative from CheckFirst, 15 May 2024, (DSA.100103, Doc ID 166-1).

2022/2065. As shown in Figure 8 below, [redacted] The representative of CheckFirst raised this issue in a meeting with the Commission,[198] and the Commission's own research confirmed similar results.[199]

[redacted]

Source: Reply to the second RFI, section VI, request 1(a), page 6 (DSA.100103, Doc ID 163-4).

(203) According to the provider of X's own Project Summary and Technical Design Document, that provider has closely monitored the provision of the funding entity through its advertisement repository through testing.[200] As of 5 April 2024,[201] the provider of X made available, in a separate section of its interface, a Google form, requiring a log in with a Google account to request the name of the person paying for a particular advertisement via credit card.[202] However, as of 14 June 2024, this form was no longer functional.[203] In any event, it cannot be considered as making the information on natural or legal person who paid for the advertisement publicly available through X's advertisement repository, as required by Article 39(2)(b) of Regulation (EU) 2022/2065. In this context, the Commission notes that the provision of reliable tools to access such information is crucial to supervision and research into emerging risks from online advertisement, including but not limited to activities that uncover coordinated disinformation or manipulative advertisements campaign that are financed by a single entity.

(204) Fourth, the search tool of X's advertisement repository is not reliable insofar as it does not provide information on commercial communications published on X and identified pursuant to Article 26(2) of Regulation (EU) 2022/2065. Until at least 25 October

197 Reply to the second RFI, section VI, request 1 (a) (DSA.100103, Doc ID 163-4).

198 Non-confidential minutes of interview with a representative from CheckFirst, 15 May 2024, (DSA.100103, Doc ID 166-1).

199 Repository outputs with missing or false information for advertisements in The Netherlands, 03 May 2024 (DSA.100103, Doc ID 134-1).

200 Reply to the second RFI, section VI, request 1 (a), page 9 (DSA.100103, Doc ID 163-6).

201 As of 05 April 2024, the form was not available: Archived X website from 5 April 2024, https://web.archive.org/web/20240405203138/https://business.x.com/en/help/ads-policies/product-policies/ads-transparency.html, accessed on 19 June 2024 (DSA.100103, Doc ID 160-53).

202 X website, https://business.x.com/en/help/ads-policies/product-policies/ads-transparency.html, accessed on 19 June 2024 (DSA.100103, Doc ID 160-33).

203 'DSA.100103' Technical Analysis, section 2.2.1 (DSA 100103, Doc ID 219-1).

2023, the repository did not provide any commercial communications.[204] Since 15 November 2023 at the latest, the repository offers a separate large .csv file that can be downloaded (see the infobox in Figure 9), but which is not searchable through the tool. This file of 13 GB, with at least 161 million entries, is artificially inflated, since every entry appears to be repeated at least 57 times, which makes it too large to be opened directly with conventional third-party tools.[205] It is therefore not reliably accessible or searchable for users.

(205) Fifth, the search tool of X's advertisement repository is not reliable, as it has serious, persistent overall issues with retrieving information through downloading the output .csv files. On 17 October 2023, WhoTargetsMe, a non-profit organisation working on the transparency of online advertising, pointed out on X that the search tool was dysfunctional[206] and hinted in response to the same post, on 19 October 2023, that the problem could be solved easily. On 19 October 2023, the provider of X was alerted in writing by CheckFirst about this dysfunctionality.[207] The Commission has confirmed that the dysfunctionality persisted until at least 27 October 2023 and concluded that it was due to an erroneous download link creation that is easy to fix with proper technical expertise.[208] Nevertheless, the outage lasted for a total of at least 10 days. Representatives from Ravineo and CheckFirst have independently further reported reliability issues in December 2023 and January 2024.[209] This evidence indicates a persistent, systemic malfunctioning of the search tool, rendering it unreliable and incompatible with the requirements of Article 39 of Regulation (EU) 2022/2065.

5.3.3. X's advertisement repository is not available through API

Figure 9: Error message from the API.

Source: ECAT (26 April 2024).

(206) In the third place, the Commission considered that X's advertisement repository does not allow queries through an API, as required in Article 39(1) of Regulation (EU) 2022/2065.

(207) In response to the second RFI,[210] the provider of X provided the Commission with public instructions on how to access the API in a separate section of X's online interface from the search in which X's advertisement repository is accessible.[211] These instructions require a potential user of X's advertisement repository to follow three steps. The first step requires the user to create a X Developer account, which includes agreeing to X's terms of service and X's Developer Agreement. The second step

---

204 Archived X website from 25 October 2024, https://web.archive.org/web/20231025225603/https://ads.twitter.com/ads-repository, accessed on 19 June 2024 (DSA.100103, Doc ID 160-45).

205 'DSA.100103' Technical Analysis, section 3.3 (DSA 100.103, Doc ID 219-1).

206 X post, https://x.com/WhoTargetsMe/status/1714276498004435112, accessed on 19 June 2024 (DSA.100103, Doc ID 160-61).

207 Correspondence between CheckFirst and X (DSA.100103, Doc ID 173-1).

208 'DSA.100103' Technical Analysis, section 2.2.2. (DSA.100103. Doc ID 219-1).

209 Non-confidential minutes of interview with a representative from CheckFirst, 15 May 2024, (DSA.100103, Doc ID 166-1).

210 Reply to the second RFI, section VI, request 1 (b), page 15 (DSA.100103, Doc ID 163-9).

211 X website, https://business.x.com/en/help/ads-policies/product-policies/ads-transparency.html, accessed on 19 June 2024 (DSA.100103, Doc ID 160-33).

involves creating an application with the developer account and generating a bearer token for authentication with the API. The third step is described as querying the API endpoint 'CreateExportReportMutation', with the generated 'bearer token' in order to request an X account of an advertiser, the Member State in which the advertisement was presented, and the time frame within which the advertisement was presented, and a report, in which the user can allegedly obtain an exportId. The status of an exportId can then be allegedly regularly queried at a status API endpoint until the report becomes downloadable. The API endpoints provided by the provider of X are not a separate API, but identical to the technical infrastructure underlying the tool with the identical limitations to search functionalities and the provision of information.[212] This means that, in practice, API access does not provide any additional elements or functionalities to receive information instead of accessing the repository via the search tool of X's advertisement repository.

(208) Repeated trials conducted by the Commission[213] following these instructions show that the last step returns the following message with error code 453 ('*You currently have access to a subset of Twitter API v2 endpoints and limited v1.1 endpoints (e.g. media post, oauth) only. If you need access to this endpoint, you may need a different access level. You can learn more here: https://developer.twitter.com/en/portal/product.*') and links to a subscription page for paid API subscription products on X (this error message is reproduced in Figure 9 above). In addition to the fact that the underlying API endpoint is identical to the endpoint that underlies the freely and publicly accessible tool described above, the Commission confirmed in trials that even an authenticated account subscribed to API access, costing USD 100 per month, cannot obtain effective access to the API, since the access tokens created with such an account still do not have the necessary access right and return the same error message.

(209) Experts consulted by the Commission on the question of API access confirmed the absence of effective access to X's API. The aforementioned representative from CheckFirst alerted the provider of X in writing, on 8 November 2023, that the API was dysfunctional without remediation or answer until at least 15 May 2024.[214] Additionally, the aforementioned representative associated with ISC-PIF contacted the provider of X in its affiliation with AI Forensics on 27 April 2024 via the email 'EU-Questions@x.com' with the aforementioned error code, but received no response.[215] The aforementioned representative of Ravineo stated that it could not access the API even after purchasing an advanced subscription to X's API plans at USD 100 per

212 'DSA.100103' Technical Analysis, section 2.3 (DSA.100103, Doc ID 219-1).

213 Trials performed on 12 April 2023, 12 May 2023, 27 January 2024, 21 February 2024, 25 March 2024, and 14 June 2024; see 'DSA.100103' Technical Analysis, section 2.3 (DSA.100103, Doc ID 219-1). Technical Analysis on Unreliability of X Ad Repository API Access (DSA.100103, Doc ID 314, 315, 316, 317, 318, 319).

214 Non-confidential minutes of interview with a representative from CheckFirst, 15 May 2024, (DSA.100103, Doc ID 166-1).

215 Non-confidential minutes of interview with an expert researcher from the Institut des Systèmes Complexes de Paris Ile de France 3, 26 April 2024 (DSA.100103, Doc ID 142-1). Email exchange between X and an expert research associated with ISC-PIF from 27 April 2024, version of 13 June 2024, (DSA.100103, Doc ID 190-1).

month.[216] According to that last representative, the provider of X did not respond to its requests to resolve the issue.[217]

(210) In sum, the malfunctioning of the API access to X's advertisement repository has been repeatedly and consistently documented, and awareness of this has been communicated to the provider of X by researchers and stakeholders since at least 8 November 2023. Therefore, the Commission took the preliminary view that the provider of X has not acted upon information regarding malfunctions concerning its repository.

(211) The Commission further notes that, in addition to the advanced subscription of USD 100 per month for API access, the provider of X offers a subscription of USD 5000 per month and an option with a customed price for API access. Even if those subscriptions were to grant more extensive access to X's advertisement repository, which the Commission has not been able to verify, the Commission considers such costs prohibitive for ordinary users and researchers. Such costs render that API access de facto ineffective for the purpose of ensuring public access to the repository to facilitate supervision and research into emerging risks.

(212) The Commission deduces that, in any event, API access for the user through this subscription is ultimately equivalent to the free access provided through the search tool of X's advertisement repository (and therefore, without the API), including all the complications described above.[218]

5.4. The arguments of the provider of X

(213) In its Reply to the Preliminary Findings, the provider of X challenged the Commission's preliminary findings regarding its non-compliance with Article 39 of Regulation (EU) 2022/2065 on the following six grounds, as set out in below.[219]

(214) In the first place, the provider of X argues that the Commission failed to provide the necessary guidance on the application of Article 39 of Regulation (EU) 2022/2065 in advance. The provider of X claims that neither Article 39 itself, nor recital 95 of that Regulation, provide sufficient information on how to comply with the obligations contained in that provision. It notes that the Commission has not issued further guidance on the structure or functionalities of the repository, as envisaged by Article 39(3) of Regulation (EU) 2022/2065, and that, therefore, providers of very large online platforms have resorted to interpreting those obligations on their own.[220]

(215) In addition, the provider of X claims that the EU legislature, in different provisions of Regulation (EU) 2022/2065, has stressed the need for additional guidance on advertising on online platforms, including in Article 39(3) and Article 46 of that Regulation.[221]

(216) In the second place, the provider of X claims that the Commission has misconstrued and misinterpreted the main items of evidence on which it relies, namely a 'Stress Test

---

216 Non-confidential minutes of interview with a representative of Ravineo, 22 April 2024 (DSA.100103, Doc ID 94-1).

217 Non-confidential minutes of interview with a representative of Ravineo, 22 April 2024 (DSA.100103, Doc ID 94-1).

218 'DSA.100103' Technical Analysis, section 2.3 (DSA. 100103, Doc ID 219-1).

219 Reply to the Preliminary Findings (DSA.100103, Doc ID 289-2).

220 Reply to the Preliminary Findings, paragraph 177 and 179.

221 Reply to the Preliminary Findings, paragraph 177.

Report' authored by the Mozilla Foundation and CheckFirst, along with corresponding meeting minutes. It claims that the finding of non-compliance with Article 39 of Regulation (EU) 2022/2065 is largely based on these two pieces of evidence.[222]

(217) The provider of X claims that the aforementioned Stress Test Report was not intended by the Mozilla Foundation and CheckFirst to be used to assess compliance with Article 39 of Regulation (EU) 2022/2065, but rather as a benchmarking exercise with the aim of promoting user friendliness and best practices in an environment characterised by significant legal uncertainty.[223] The provider of X bases its claims on draft minutes and a draft Stress-test report that were not cited in the Preliminary Findings, but were made available to it through the data room procedure.[224]

(218) In the third place, the provider of X claims that external independent auditors have recently concluded that '*X complies with the obligation to compile and make publicly available a repository of information about advertisements presented on its online interfaces*'. The provider of X notes that the independent auditors reached a positive conclusion regarding the obligations of Article 39 of Regulation (EU) 2022/2065.[225]

(219) In the fourth place, the provider of X claims that the search tool of X's advertisement repository is searchable.

(220) First of all, it argues that the Commission has neglected to define or provide guidance on the terms 'searchable', 'multicriteria queries', how to comply with the intended purpose of the provision, or when the delivery of a report is considered excessively slow pursuant to Article 39 of Regulation (EU) 2022/2065.[226]

(221) In addition, it claims that '*X (i) does allow users to search stored advertisements through multiple search criteria (including the name of the advertiser, the Member State where the ads were served and the date range within which the advertisement was presented), (ii) does display results in a .csv file that enables searches, filtering and complex data manipulation, and (iii) has made an effort to gradually reduce the response time to the minimum possible in light of existing technical constraints*.'[227]

(222) First, the provider of X claims that the search tool allows multicriteria queries to access X's advertisement repository. As regards multicriteria queries, the provider of X claims that the Commission's interpretation of Article 39 of Regulation (EU) 2022/2065 would appear to require the repository to provide at least one field of search for each of the items of information listed in Article 39(2) of that Regulation.[228] It claims that the key cause of the Commission's confusion is the attempt to incorrectly conflate the two paragraphs of Article 39, and that this interpretation has no apparent legal basis.[229] It argues that each paragraph of that provision governs difference aspects of the advertisement repository.

(223) On one hand, the provider of X claims that Article 39(1) of Regulation (EU) 2022/2065 is concerned with the functionality of the advertising repository and requires a 'searchable and reliable tool' with particular capacities, such as multicriteria

222 Reply to the Preliminary Findings, paragraph 190.
223 Reply to the Preliminary Findings, paragraph 191.
224 The Provider of X's claims regarding accessibility of such documents are discussed in section 7.
225 Reply to the Preliminary Findings, paragraphs 187 and 211.
226 Reply to the Preliminary Findings, paragraph 195.
227 Reply to the Preliminary Findings, paragraph 196.
228 Reply to the Preliminary Findings, paragraph 197.
229 Reply to the Preliminary Findings, paragraph 198.

queries. The emphasis is on how the repository operates, and the term 'searchable', relates to the system's ability to retrieve information based on multiple, undefined search criteria.[230] On the other hand, the provider of X claims that Article 39(2) of Regulation (EU) 2022/2065 is concerned with the substantive content that must be made available through the repository. According to the provider of X, the wording '*containing the information referred to in paragraph (2) in Article 39 of Regulation (EU) 2022/2065'* supports this proposed interpretation.[231]

(224) The provider of X further claims that the Commission's interpretation of Article 39 of Regulation (EU) 2022/2065 has no apparent support in academic literature,[232] the Stress Test conducted by the Mozilla foundation and CheckFirst, or from other independent guidelines or initiatives.[233]

(225) Finally, the provider of X claims that all other very large online platforms have interpreted Article 39 of Regulation (EU) 2022/2065 in a similar way to it, and that the central concern of Article 39 is about fulfilling the functionality requirements.[234]

(226) Second, the provider of X claims that the display of search results in a .csv file complies with Article 39 of Regulation (EU) 2022/2065, including the intended purpose of that provision. According to the provider of X, Regulation (EU) 2022/2065 does not require that search results are presented in a specific section of X's online interface. It claims that Article 39 of Regulation (EU) 2022/2065 necessitates the repository to be on an online interface, but that provision does not require the repository to be a 'search tool'. The search tool itself needs to be publicly available, rather than the results.[235]

(227) The provider of X further claims that Regulation (EU) 2022/2065 does not disallow it from displaying search results in a .csv file. The word 'compile' in Article 39 of Regulation (EU) 2022/2065 implies that a structured format should be used. Absent a technical standard specified by the Commission, X has chosen a file that is compatible with X's software.[236] The provider of X claims that a .csv file allows for control over the document and can be integrated into a data analysis tool allowing for data manipulation. It states that utilising a .csv file to display results is more secure against cyber-attacks and reduces risk of unauthorized access to potentially sensitive data.[237] The provider of X specifically references in this respect '*A Standard for Universal Digital Ad Transparency*'[238] and '*Access to Data and Algorithms: For an Effective*

230 Reply to the Preliminary Findings, paragraph 198.
231 Reply to the Preliminary Findings, paragraph 198.
232 Reply to the Preliminary Findings, footnote 226.
233 Reply to the Preliminary Findings, paragraph 199.
234 Reply to the Preliminary Findings, paragraph 200.
235 Reply to the Preliminary Findings, paragraph 202.
236 Reply to the Preliminary Findings, paragraph 202.
237 Reply to the Preliminary Findings, paragraph 202: Users and researchers can determine who can access it, the storage, distribution and customization of the file. Online formats may be subject to the management of a provider, whereas a .csv file can be accessed independently, allowing for long-term persistence. This kind of file allows users to combine data sets and apply various filters. A .csv files can be accessed without an internet connection, and are secure against cyber-attacks, limiting the risk of unauthorized access.
238 Edelson, Laura, Jason Chuang, Erika F. Fowler, Michael M. Franz and Travis Ridout 'A Standard for Universal Digital Ad Transparency', Knight First Amendment Institute at Columbia University, (December 2021, https://knightcolumbia.org/content/a-safe-harbor-for-platform-research Accessed 29 January 2025. DSA.100103, Doc-ID 306.

DMA and DSA Implementation'[239] as academic references supporting the use of .csv files. As regards the Commission's finding that opening a .csv file requires the use of third-party software, the provider of X notes that accessing X's advertisement repository through an online interface also requires third party software, namely an internet browser.[240]

(228) Third, the provider of X claims that the response time to obtain a report from X's advertisement repository does not affect the searchability of the search tool and that it has made an effort to gradually reduce the response time to the minimum possible in light of existing technical constraints. It also claims that is has not made a deliberate and conscious decision to slow the response time of the search tool, and that there is no evidence to support this claim.[241] Finally, the provider of X claims that altering the basic IT architecture to have a near-instant response time for results would be costly, time consuming, with no guarantees of success and lead to possible inaccuracies.[242]

(229) In the fifth place, the provider of X claims that its advertisement repository is reliable.

(230) First, the provider of X claims that it has made reasonable efforts to ensure the reliability of the search tool and the API. It argues that the notion of 'reliability' used in Article 39(1) of Regulation (EU) 2022/2065 should be interpreted as an obligation applicable to the implementation process and the tools to meet regulatory requirements, rather than achieving a specific outcome.[243] It further contends that it has been making continuous engineering efforts to ensure reliability, accuracy, and completeness of its advertising repository.[244] It also states that it has made improvements by including additional data. According to the provider of X, all these efforts exceed the reasonable efforts required of providers under Article 39 of Regulation (EU) 2022/2065.[245] Second, the provider of X contends that its advertisement repository is complete and contains information on all the advertisements displayed on X's online interface. It claims that the Commission relied on a particular research study, referencing the ISC PIF report, that uses a 'very limited sample'. According to the provider of X, the aforementioned report is based on 151 users and 250 advertisements, and it solely concerns advertisements and users in France. Additionally, the provider of X states that this study was conducted at the beginning of the entry into force of Regulation (EU) 2022/2065 (between 1 September 2023 and 18 April 2024).[246]

---

239 Edelson, Laura, Inge Graef, and Filippo Lancieri 'Access to Data and Algorithms: For an Effective DMA and DSA Implementation' Centre on Regulation in Europe (CERRE), page 50, March 2023, DSA.100103, Doc ID 395, Accessed 6 May 2025. DSA.100103, Doc-ID 305.

240 Reply to the Preliminary Findings, paragraph 202.

241 Reply to the Preliminary Findings, paragraph 205. The provider of X claims that tool and the API function are determined by X's technical architecture. This operates on [redacted] It is noted that the response time depends on the users' internet connection.

242 Reply to the Preliminary Findings, paragraph 206.

243 Reply to the Preliminary Findings, paragraphs 210, 211.

244 Reply to the Preliminary Findings, paragraph 211.

245 Reply to the Preliminary Findings, paragraph 212.

246 Reply to the Preliminary Findings, paragraph 213.

(231) The provider of X further claims that the Commission's research is focused on punctual malfunctions, rather than on evidence that could prove the accuracy of its search tool. The provider of X also contends that the Commission disregards a part of the Mozilla Foundation study invoked in the Preliminary Findings that confirms the accuracy of X's repository.[247] The provider of X claims that meeting minutes indicate that Ravineo also confirmed the completeness of the X's advertisement repository.[248]

(232) The provider of X then claims that Article 39 of Regulation (EU) 2022/2065 does not prevent advertising content being accessed through a URL in a .csv file, rather displaying it on a user interface. It further asserts that this allows users themselves to access and verify the information provided.[249] According to the provider of X, displaying a URL is a practical solution for several reasons. In the alternative, compiling and hosting advertising content in a repository would require '*a disproportionate amount of resources for storage and processing power'*, resulting in higher operational costs and a greater environmental impact.[250] The provider of X claims that this method of complying with Article 39 of Regulation (EU) 2022/2065 allows for the direct supervision and research of systemic risks, in line with recital 95 of that Regulation. It further argues that it does not hinder any research by providing a URL, but allows researchers to access a primary source.[251]

(233) Finally, the provider of X claims that if a user deletes or alters an advertisement, this is beyond its control, and Regulation (EU) 2022/2065 does not require from the provider of X to retain content if an advertiser has chosen to remove it on the online platform. The provider of X further claims that providers of very large online platforms must respect the advertisers' rights to delete content, in absence of an explicit requirement to the contrary.[252]

(234) Third, the provider of X argues that its advertisement repository allows users to identify the natural or legal person who paid for and on whose behalf the advertisement is presented. It further contends that it has made the legitimate choice of differentiating between the advertiser who paid for presentation of the advertisement via insertion order or credit line and those who paid [redacted].[253] The provider of X claims that due to privacy concerns, the only solution to avoid undermining fundamental rights is to allow advertisers to request the name of the person paying for the advertisement through a Google form.[254] Finally, the provider of X argues that a failure to respect the rights of third parties could negatively impact X's business interests and relationships with advertisers.[255] Without a clear legal obligation to disclose the identity of those persons, it claims not to have a sufficient legal basis to interfere with the privacy rights of third parties paying [redacted] and it cannot disclose their identity immediately.[256]

---

247 Reply to the Preliminary Findings, paragraph 214.
248 Reply to the Preliminary Findings, paragraph 215.
249 Reply to the Preliminary Findings, paragraph 217.
250 Reply to the Preliminary Findings, paragraph 218.
251 Reply to the Preliminary Findings, paragraph 223.
252 Reply to the Preliminary Findings, paragraph 219.
253 Reply to the Preliminary Findings, paragraph 223.
254 Reply to the Preliminary Findings, footnote 254. It references judgment of 22 November 2022, WM and Sovim SA v Luxembourg Business Registers, C-37/20 and C-601/20, EU:C:2022:912, concerning the interference with fundamental rights.
255 Reply to the Preliminary Findings, paragraph 224.
256 Reply to the Preliminary Findings, paragraph 225.

(235) Fourth, the provider of X claims that information on commercial communications is easily accessible via the hyperlink 'Commercial Communications Disclosure' available on X's advertisement repository since November 2023. This results in the creation of a separate and completely searchable .csv file allowing researchers to filter and complete their own analysis.[257] The provider of X argues that the Commission never previously raised concerns about the size of the Commercial Communications report to X.[258]

(236) Fifth, the provider of X claims there are no serious or systemic issues compromising the reliability of the search tool of X's advertisement repository. It claims that the Commission's allegations regarding the reliability of that tool are based on isolated incidents, and that the Commission has failed to adduce evidence of persistent and systemic malfunctioning.[259]

(237) In the sixth place, the provider of X claims that X's advertisement repository can be accessed through an API, as required by Article 39 of Regulation (EU) 2022/2065. The provider of X claims that it has made its advertising repository available via its user interface (i.e. search tool) and an API, which relies on the same architecture, and that both give access to the same information.[260] The provider of X stresses that access to X's advertisement repository through the API is free of charge, available for all users,[261] and does not require a paid subscription.[262]

(238) The provider of X claims that *'[t]he sole reason why the Commission was unable to access the API's endpoint during the trials it conducted is because of a bug that was fixed as soon as X identified the issue and its cause (on 13 July 2024).*'[263] The provider of X claims that, contrary to what the Commission claims, the malfunctioning of X's public API has not been repeatedly and consistently documented and that the Commission's claim is based on solely on the Commission's trials and '*on the experience of the same two researchers on whose opinions the Commission bases all its allegations against X in relation to Article 39 DSA. Labelling the API as 'ineffective' and 'malfunctioning' based on such scant and insubstantial evidence is unwarranted and disproportionate*'.[264] The provider of X also claims that X has not been made repeatedly aware of any alleged dysfunctionalities by researchers and other stakeholders.[265]

5.5. The Commission's assessment of the provider of X's arguments

(239) For the reasons set out in the following subsections, the Commission finds that the arguments put forward by the provider of X are unable to call into question the Commission's finding of an infringement of Article 39 of Regulation (EU) 2022/2065 by that provider, as set out in section 5.3 above.

257 Reply to the Preliminary Findings, paragraph 226.
258 Reply to the Preliminary Findings, paragraph 227.
259 Reply to the Preliminary Findings, paragraph 230.
260 Reply to the Preliminary Findings, paragraph 232.
261 Reply to the Preliminary Findings, paragraph 233.
262 Reply to the Preliminary Findings, paragraph 234 and 235.
263 Reply to the Preliminary Findings, paragraph 236.
264 Reply to the Preliminary Findings, paragraph 237.
265 Reply to the Preliminary Findings, paragraph 238.

5.5.1. The claim that the Commission has failed to provide guidance

(240) In the first place, the Commission considers that the provider of X's claim of an alleged failure on the part of the Commission to provide guidance in accordance with Article 39(3) of Regulation (EU) 2022/2065 is ineffective. Article 39(3) of Regulation (EU) 2022/2065 allows but does not oblige the Commission to issue guidelines on structure, organisation, and functionalities of the repositories referred to in Article 39 of Regulation (EU) 2022/2065. Article 39 of that Regulation is, by its very terms, self-executing and directly applicable to providers of very large online platforms and of very large online search engines.

(241) The Commission considers that, were it required to issue guidance before taking any enforcement steps (*quod non*), it would be effectively prevented from enforcing obligations laid down by Regulation (EU) 2022/2065 for a substantial period of time, leading to a delay in the full achievement of the objectives that that Regulation aims to achieve. This would give rise to a risk of potentially allowing an online environment which threatens the fundamental rights provided for in the Charter to persist and develop.[266] In this regard, the Commission notes that the Union legislature attached particular importance to applying Regulation (EU) 2022/2065 as quickly as possible to providers of very large online platforms and of very large online search engines, as demonstrated by the advanced entry into force of that Regulation for those providers.[267]

5.5.2. The claim that the Commission allegedly misconstrued and misinterpreted main items of evidence on which it relies

(242) In the second place, as regards the provider of X's claim that the Commission allegedly misconstrued and misinterpreted the main items of evidence on which it relied to find an infringement of Article 39 of Regulation (EU) 2022/2065, it is based on the misconception that the Commission relied only, or even predominantly, on the Stress Test Report authored by the Mozilla Foundation and CheckFirst, along with corresponding meeting minutes, to substantiate that finding. As is clear from the Preliminary Findings, the Commission adduced an overall body of evidence to establish its finding of non-compliance with that provision, most importantly the trials conducted by the Commission demonstrating the shortcomings of X's advertisement repository.[268] Therefore, without undermining the importance of the Stress Test Report and the corresponding minutes, the provider of X's characterisations of these reports and those minutes as '*main items of evidence*', on which the Commission's assessment was '*largely based*', is misleading.

(243) The provider of X's claim also appears to be founded on the incorrect premise that, for the Commission to rely on a study conducted by a third party in its assessment of a provider's compliance with Regulation (EU) 2022/2065, that study must have the explicit objective to assess compliance with that Regulation. That is not the case as the Commission is able to rely on any pertinent item of evidence from which it may conclude that the provider has not complied with its obligations under Regulation

[266] Order of the Vice President of the Court 24 March 2024, Amazon Services Europe v Commission, C-639/23, ECLI:EU:C:2024:277, paragraph 157.

[267] Order of the Vice President of the Court, Amazon Services Europe v Commission, C-639/23, ECLI:EU:C:2024:277, paragraph 162.

[268] 'DSA.100103' Technical Analysis (DSA 100.103, Doc ID 219-1).

(EU) 2022/2065, even if the purpose of that evidence was not to establish such non-compliance.

(244) In any event, the provider of X is wrong to allege that the Commission drafted the minutes anything but faithfully to reflect what was discussed during the meeting. Earlier drafts of those minutes do not show otherwise. As the provider of X's specified external counsel was able to establish on the basis of the documents provided to it through the data room procedure, Mozilla Foundation requested five changes to be made to the minutes in light of its explicit desire to avoid the impression that the study was conducted with the aim of carrying out an assessment of X's compliance with Regulation (EU) 2022/2065. These changes did not in any way impact the substance of Mozilla Foundation's findings or contest the factual correctness of the minutes. The final non-confidential version of the minutes was approved by the Mozilla Foundation and CheckFirst and it is the only version considered in the Commission's investigation.

(245) The provider of X also misrepresents the Stress Test report when it claims there was no adherence to any legislative requirement. According to the meeting minutes, the representative of CheckFirst explained that '*the idea for the methodology was to perform twelve different tests, against the background of the DSA and Mozilla's principles mentioned before, for instance related to accuracy, reliability, and completeness*'.[269] Furthermore, the report states that the *'evaluation is based on a combination of the DSA's requirements for VLOPS in Article 39(1-2) and five guidelines written by Mozilla and other experts in 2019*'.[270] (emphasis added)

5.5.3. The claim regarding the conclusions of the independent audit

(246) In the third place, as regards the provider of X's claim that it received a positive conclusion in the audit conducted under Article 37 of Regulation (EU) 2022/2065 in relation to its compliance with Article 39 of that Regulation, such a positive conclusion does not prejudge of the provider's compliance nor precludes the Commission from exercising its supervisory and enforcement powers including in relation to that provision and from concluding that the provider of X has failed to comply with its obligations under that Regulation, also taking into account that its finding of non-compliance in the present case is based on a body of evidence. The purpose of the annual audit is to contribute and inform the Commission in its supervision and enforcement activities under Regulation (EU) 2022/2065. However, this does not mean that the Commission must accept at face-value the outcome of such an audit.[271]

(247) In any event, the Commission observes that according to the Audit Report 2024, the auditors were unable to access X's advertising repository through the API, despite the clear obligation in Article 39 of Regulation (EU) 2022/2065 to make the repository publicly available through API. This finding alone points to a lack of compliance with

269 Non-confidential minutes of interview with a representative from CheckFirst, 15 May 2024, (DSA.100103, Doc ID 166-1).

270 Mozilla Foundation and CheckFirst, Full Disclosure: Stress testing tech platforms' ad repositories, 16 April 2024 (DSA.100103, Doc ID 171-1).

271 The Audit Report 2025 does not include conclusions on obligations that are covered by the Commission's proceedings, see the Audit Report 2025, page 5 (DSA.100101, Doc ID 289-2; DSA.100102, Doc ID 382-2; DSA.100103, Doc ID 325-2).

that Regulation. This deficiency of X's advertising repository is further assessed in the section 5.5.4. below. [272]

5.5.4. The claim that the search tool is searchable

(248) In the fourth place, as regards the claim that the search tool of X's advertisement repository is searchable, the Commission finds that that claim is unfounded for the following reasons.

(249) First, the Commission notes that Article 39(1) of Regulation (EU) 2022/2065 obliges providers of very large online platforms and of very large online search engines to make a repository that contains specific information about advertisements available through a *'searchable'* tool that allows *'multicriteria queries'*. The Commission observes that, by imposing this requirement, the Union legislature has sought to enable public and regulatory scrutiny over vast amounts of advertisements that are presented on such platforms and search engines, including in relation to the dissemination of illegal advertisements.[273] The Commission considers that the requirement to provide a searchable tool that allows multicriteria queries to enable public access to advertisements should be construed in light of this objective.

(250) The provider of X appears to argue that it complies with this requirement by providing a tool that allows searching for a single advertiser in a single Member State at a particular moment in time. This limitation means that most of the information that is required to be made available through the tool is not searchable through a tool allowing multicriteria queries within the meaning of Article 39(1) of Regulation (EU) 2022/2064.

(251) The Commission takes the position that the provider of X, by providing such a tool, has failed to comply with the requirement to provide a *'searchable […] tool that allows multicriteria queries'* within the meaning of Article 39(1) of Regulation (EU) 2022/2065. The Commission considers that, were it to accept the provider of X's interpretation, the obligation to make information about advertisements available through a *'searchable […] tool that allows multicriteria queries'* would be rendered ineffective in achieving its express objective of enabling public and regulatory scrutiny on the risks stemming from online advertising.

(252) The requirement of a 'searchable tool' allowing for 'multicriteria queries' laid down in that provision serves a twofold purpose. On one hand, it facilitates broader research into societal risks stemming from online advertising. Recital 95 of Regulation (EU) 2022/2065 clarifies the EU legislature's intention to facilitate such research by recognising the risks that advertising systems used by very large online platforms and very large online search engines can pose. Such advertising systems require further public and regulatory supervision, which should be ensured through public access to an advertising repository to facilitate the necessary research into emerging risks.

(253) On the other hand, it may work to aid the competent authorities in supervising compliance with Regulation (EU) 2022/2065. The repository may, for example,

---

272 '*An API function was not tested successfully. The documentation provided on the steps to query the API is not very clear, and during tests to validate the functionality of the API itself, an error was generated stating that the account did not have access to the reporting despite owning all account tiers as per DSA requirement and the X platform accounts page.*', the Audit Report 2024, page 255 (DSA.100103, Doc ID 293).

273 Recital 95 of Regulation (EU) 2022/2065.

inform the assessment of compliance with other provisions of that Regulation, including those not contained in Section 5 of Chapter III of the Regulation. Recital 68 of the Regulation clarifies that online advertising can contribute to significant risks, ranging from advertisements that are illegal to publish, that support financial incentives for publication, or that amplify illegal or harmful content, such as scam advertisements. In addition, Recital 95 provides for several examples of other risks as a result of online advertising, including manipulative techniques, disinformation, and the delivery of advertisements to persons in vulnerable situations, such as minors.

(254) Second, as regards the provider of X's claim that the different paragraphs of Article 39 of Regulation (EU) 2022/2065 govern different aspects of the advertisement repository, the Commission considers that this interpretation misconstrues the structure of that provision. Article 39(1) of Regulation (EU) 2022/2065 lays down the obligation to establish and '*make publicly available […] a repository containing the information referred to in paragraph 2*'. In addition, it specifies how providers of very large online platforms and search engines should provide access to that repository. Article 39(2) of Regulation (EU) 2022/2065 specifies the requirement laid down in the first paragraph, by detailing the (minimum) information that the repository must provide concerning the advertisements displayed on the very large online platform or online search engine. Therefore, instead of governing different aspects, Article 39(1) and Article 39(2) of Regulation (EU) 2022/2065 are directly interconnected.[274]

(255) As explained in recitals 249 and 250 above, the provider of X's flawed interpretation of the notion of 'multicriteria queries' in designing its tool prevents that tool from achieving the public and regulatory scrutiny which the Union legislature set out to achieve through the enactment of Article 39 of Regulation (EU) 2022/2065.

(256) The Commission observes in this regard that Article 39 of Regulation (EU) 2022/2065 does not merely require the advertisement repository to be made available through 'a reliable **search** tool', as the provider of X suggests, but rather through a '**searchable** and **reliable** tool'. Where certain information related to advertisements referred to in Article 39(2) of that Regulation is not directly searchable through the tool, the provider has failed to make the repository containing this information '*publicly available…through a searchable tool*' as required by Article 39(1) of the Regulation.

(257) As stated in recital 182 above, the search tool of X's advertisement repository only allows most of the information related to advertisements listed in Article 39(2) of Regulation (EU) 2022/2065 to be obtained indirectly by searching a single advertiser in a single Member State at a particular moment in time. This limitation means that most of the information that the provider of X is required to be made available through the search tool is not searchable nor searching such information is possible through multi-criteria queries. This also hinders research or supervision into emerging risks that stretches beyond a single advertiser in more than one Member State. Therefore, the provider of X has taken an overly restrictive interpretation of Article 39 of Regulation (EU) 2022/2065 which is not in line with the objective of that provision.

[274] This is further corroborated by the fact that Article 39(1) of Regulation (EU) 2022/2065 contains several additional obligations in relation to the information that must be included in the repository, including that the repository shall not contain any personal data of the recipients of the service to whom the advertisement was or could have been presented and that the provider of the very large online platform shall make reasonable efforts to ensure that the information is accurate and complete.

(258) As regards the provider of X's claim that other advertisement repositories contain the same or similar limited number of search criteria, the Commission recalls, leaving aside that every case must be assessed on its own merits, that an operator cannot claim that it has complied with its obligations by pointing to other operators that have failed to comply with their obligations.[275] In any event, the Commission notes that it has initiated investigations into the compliance of two other providers of very large online platforms with Article 39 of Regulation (EU) 2022/2065 and addressed requests for information to several other providers.[276]

(259) Third, as regards the provider of X's claim that that the display of search results in a .csv file complies with Article 39(1) of Regulation (EU) 2022/2065, the Commission observes that that provision requires that '*a repository containing the information referred to in paragraph 2*' shall be made '*publicly available in a specific section of [the provider's] online interface, through a searchable and reliable tool*'. (emphasis added) This means that relevant providers bound by this obligation are required to make publicly available an advertising repository through a tool that is searchable and that, consequently, has search results that can be viewed within the interface. Furthermore, the provision specifically refers to a 'searchable […] tool', not a 'search tool' as the provider of X suggests.

(260) Similarly, the provider of X's claim concerning platform independence for opening a .csv file is irrelevant, given that Article 39 of Regulation (EU) 2022/2065 requires the provider to make publicly available an advertisement repository that includes information that must be provided on an online interface.

(261) As regards the provider of X's claim that the use of third party software to access the .csv file does not amount to non-compliance with Article 39 of Regulation (EU) 2022/2065 by pointing to the fact that a user must use a third-party browser to access the advertisement repository, the Commission observes that the reference to an 'online interface' in that provision already presupposes the use of software that has the capability to display such an interface. The need for additional third-party software to open a .csv file constitutes an unnecessary barrier and requires special technical expertise, as opposed to a situation where the results are available on an online interface accessible via an internet browser.

(262) The Commission considers, in this regard, the provider of X's reference to 'A Standard for Universal Digital Ad Transparency'[277] as supporting the use of .csv files

---

275 Judgment of 31 May 2018, Groningen Seaports v. Commission, Case T-160/16, EU:T:2018:317, paragraph 116.

276 The proceedings initiated under the DSA against AliExpress, TikTok and X related to alleged infringements to Article 39. See European Commission, 'Commission opens formal proceedings against TikTok under the Digital Services Act', available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_24_1485 See also European Commission, ' Commission opens formal proceedings against AliExpress under the Digital Services Act', available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_24_1485.
The Commission addressed requests for information to Amazon (https://digital-strategy.ec.europa.eu/en/news/commission-requests-information-amazon-under-digital-services-act), Pornhub, Stripchat, XVideos (https://digital-strategy.ec.europa.eu/en/news/commission-requests-information-under-digital-services-act-pornhub-stripchat-and-xvideos-their), Apple App store, Booking.com, Google designated services and Bing (https://digital-strategy.ec.europa.eu/en/news/commission-requests-information-under-digital-services-act-apple-bookingcom-google-and-microsoft).

277 Edelson, Laura, Jason Chuang, Erika F. Fowler, Michael M. Franz and Travis Ridout 'A Standard for Universal Digital Ad Transparency', Knight First Amendment Institute at Columbia University, (

to be irrelevant. That particular advertising standard does not consider the obligation in Article 39 of Regulation (EU) 2022/2065 that the advertisement repository must be made publicly available through the online interface. Rather, that standard notes that for the large volume of ads that it suggests being included, a .csv file might not be an appropriate method for displaying results.

(263) As to the provider of X's reference to 'Access to Data and Algorithms: For an Effective DMA and DSA Implementation', it does not explicitly mention the use of .csv files in any way. As regards advertisement repositories on an online interface, this document specifically recommends the use of a web portal. 'In practical terms, this means the development of both an API with specific characteristics for bulk access and a searchable web portal for ad-hoc access by less technically sophisticated parties…The DSA's overall purpose of increasing transparency over which ads are displayed, how they are targeted and whether there are any hidden risks in this targeting process is better served by a mechanism such as a web portal that is accessible to a wider range of European citizens'(emphasis added).[278]

(264) The same concern was highlight in X's Audit Report for 2024, in which the independent auditors note that 'the search output of this tool is overly challenging for the average user, as the specificity required in the search string to navigate the volume of this output is too complex and the alternative download option is increasingly impractical' (emphasis added).[279]

(265) Fourth, as regards the response time needed to obtain a report from X's advertisement repository, the Commission considers that there is no support for the claim by the provider of X that X's technical architecture prevents faster response times to obtain search results from X's advertising repository through the search tool. As explained in recital 189 above, the Commission's accelerated procedure to retrieve results from the provider of X's advertising repository demonstrated that reports can be delivered in 15 seconds rather than 200 seconds, and potentially faster, without any inaccuracies or other issues.

(266) In addition, the provider of X has confirmed that the technical system ('endpoints'), from which reports are downloaded, was functioning in 100 percent of cases.[280] This includes the period when the Commission obtained numerous reports while circumventing the intentional delay of 200 seconds and accelerating report creation to 15 seconds.[281] The accelerated delivery time did not create any inaccuracies or other issues to the system, which indicates that there should be no technical limitations to provide the reports much faster than in 200 seconds.

(267) More important, the Commission notes that the provider of X is bound by the obligation to make the advertisement repository publicly available through a searchable and reliable tool, as required in Article 39(1) of Regulation (EU) 2022/2065. This means that any technical architecture that the provider of the very large online platform or very large search engine in question puts in place to comply

---

December 2021, https://knightcolumbia.org/content/a-safe-harbor-for-platform-research Accessed 29 January 2025. DSA.100103, Doc-ID 306

278 Edelson, Laura, Inge Graef and Filippo Lancieri, 'Access to Data and Algorithms: For an Effective DMA and DSA Implementation' Centre on Regulation in Europe (CERRE), page 50, March 2023, Accessed 6 May 2025, DSA.100103, Doc-ID 305

279 Audit Report 2024, section D.1, page 253 (DSA.100103, Doc ID 293).

280 Reply to the Preliminary Findings, footnote 207.

281 'DSA.100102' Technical Analysis (DSA.100103, Doc ID 219-1).

with such obligation has to be designed in such a way that compliance is *de facto* achievable, i.e., any possible limitations in that architecture are not a valid reason to excuse non-compliance with Article 39(1) of Regulation (EU) 2022/2065. If the architecture chosen by the provider hinders compliance with the legal obligation, it is for that provider, to implement the necessary changes. The legal obligations of Regulation (EU) 2022/2065 cannot be re-interpreted by the provider of a platform to fit the current technical architecture of the advertising repository.

(268) As a final point, the Commission clarifies that, as regards the performance of multicriteria searches through the tool, multicriteria search queries need to cover the categories in Article 39(2) of Regulation (EU) 2022/2065 in relation to advertisements, but do not need to cover commercial communications identified pursuant to Article 26(2), as referred to in Article 39(2)(f) of that Regulation. The Commission considers that this is the case because the information referred to in Article 39(2)(f) of that Regulation is, given its nature, not amenable to be included in a searchable tool allowing multicriteria queries designed to make information about advertisements available. Therefore, contrary to the preliminary finding presented in recital 181 above, the Commission considers that the exclusion of such commercial communications as a search option to perform multicriteria searches is not relevant for the purpose of establishing an infringement of Article 39 of Regulation (EU) 2022/2065. This however does not exempt the provider of X from the other obligations under Article 39 of that Regulation.

5.5.5. The claim that the search tool is reliable

(269) In the fifth place, the Commission rejects the provider of X's claims that the search tool is reliable on the following grounds.

(270) First, as regards the provider of X's claim that it has made reasonable efforts to ensure the reliability of the search tool, the Commission has already observed that Article 39 of Regulation (EU) 2022/2065 only refers to reasonable efforts on the part of the provider when it comes to ensuring *'that the information is accurate and complete*'. This simply means that the provider must make reasonable efforts to collect accurate and complete information from the advertiser. In contrast, no mention is made of 'reasonable efforts' in Article 39 of Regulation (EU) 2022/2065 when it comes to creating a 'reliable' tool to make the repository publicly available.

(271) Second, as regards the provider of X's claim that its advertisement repository is complete and contains information on all the advertisements displayed on X's interface, the Commission observes that Article 39(1) of Regulation (EU) 2022/2065 requires providers of online platforms to make certain information available '*for the entire period during which they present an advertisement and until one year after the advertisement was presented for the last time on their online interfaces'.* Pursuant to Article 39(1) of Regulation (EU) 2022/2065 this should contain the information listed in paragraph (2) of that Article, including *'the content of the advertisement*'. As explained in recitals 195 to 198 above, X's advertisement repository does not include information for a year after the advertisement was presented, nor does it contain information on the content of the advertisement.

(272) By requiring providers to make information available '*until one year after the advertisement was presented for the last time*', the EU legislature explicitly opted for the retention of such information after the advertisement is no longer being served and after it has been removed. This requirement ensures, for example, that bad faith actors are prevented from frustrating any supervision or research into, for example,

potentially illegal advertisements by removing their advertisements. Article 39(3) of Regulation (EU) 2022/2065 only contains a narrowly tailored exemption for the retention of certain information on advertisements that were removed or to which access was disabled based on alleged illegality or incompatibility with the provider's terms which is not relevant in this instance.[282]

(273) The Commission further notes that the provider of X already hosts all the advertising content required for the repository, namely for the very purpose of serving advertisements to recipients of the service, and finds that it would not require '*a disproportionate amount of resources for storage and processing power*' to include that information in X's advertisement repository, rather than a URL to that information. Therefore, providing the content of the advertisements in the repository that has already been presented cannot reasonably be held to impose an additional burden on the provider of X.

(274) Finally, the Commission considers the provider of X misrepresents the study conducted by a researcher at ISC-PIF. That report is based on a total of 31 362 advertisements presented to 151 X users based in France between 1 September 2023 to 18 April 2024.[283] The study used only advertisements that were served in France to reflect how X's tool only allows for the search of one Member State at a time. However, due to the lack of searchability and API access of the repository, the researcher was not able to retrieve all 31 362 advertisements that were served originally on X.[284] As a result, two subsamples were studied, the first one concerning 12 175 unique advertisements that were placed by 20 of the 200 largest advertisers in France. The researcher identified that the advertisements of one particular advertiser were not present in the repository search results at all, which narrowed the potential pool of advertisements to appear in the search results to 11 045. Ultimately, the researcher found that only 5 768 advertisements (i.e. 52.2%) were in the results. The study considered a second 'cross section' that consisted of 250 randomly drawn advertisements from 165 unique advertisers shown to the 151 users from the original total of 31 362 advertisements. In this 'cross section' sample, which is representative of the advertisements shown to the 151 users, ISC-PIF found that only 58% of ads were included in the repository search results. In conclusion, these samples demonstrate that roughly half of the advertisements that should have been included in X's advertisement repository pursuant to Article 39 of Regulation (EU) 2022/2065 were missing from the results. The provider of X only refers to one of the sub-samples of the study and neglects to mention the other sample which contributed to the Commission's preliminary findings.

(275) In addition, the provider of X claims that the Mozilla Foundation study confirms the accuracy of X's advertising repository and that the Ravineo meeting minutes indicate that Ravineo confirmed the completeness of X's advertisement repository, but that the Commission instead chose to rely on a study illustrating issues with X's repository. The Commission takes the view that the aforementioned study conducted by ISC-PIF is not comparable to the evidence from Mozilla Foundation and CheckFirst, along

282 See Article 39(3) of Regulation (EU) 2022/2065.

283 Bouchaud Paul, X Ads Repository – Evaluating X Ads Repository (In)Completeness Under Article 39 of the Digital Services Act, 8 May 2024 (DSA.100103, Doc ID 100-1) page 4.

284 Bouchaud Paul, X Ads Repository – Evaluating X Ads Repository (In)Completeness Under Article 39 of the Digital Services Act, 8 May 2024 (DSA.100103, Doc ID 100-1).

with the Ravineo meeting minutes. This is due to the differences in the study methodology and large sample size used in the ISC-PIF study.

(276) Third, as regards the provider of X's claim that it is cannot disclose the identity of natural persons who paid for or on whose behalf an advertisement is presented on X without infringing that person's right to privacy, the Commission observes that Article 39(2)(c) of Regulation (EU) 2022/2065 requires the repository to include 'the natural or legal person who paid for the advertisement', where that person is different from the natural or legal person on whose behalf the advertisement is presented.

(277) The Commission further observes that disclosure of the name of the natural person on whose behalf the advertisement is presented is already required under existing Union law, in particular under Article 6(b) of Directive (EU) 2000/31[285] and Article 26(1)(b) of Regulation (EU) 2022/2065.[286] Moreover, the Commission additionally notes that the disclosure of the natural or legal person who paid for the advertisement, if that person is different from the natural or legal person on whose behalf the advertisement is presented, is also required under Article 26(1)(c) of Regulation (EU) 2022/2065.

(278) The Commission also observes that the very argument regarding the protection of the privacy of natural persons raised by the provider of X has already been addressed and rejected by the President of the General Court in his orders in *Aylo Freesites v Commission* and *WebGroup Czech Republic v Commission*.[287] In the *Aylo Freesites* order, the President of the General Court found that '*the mere fact of being listed in the advertisement repository as a natural person does not necessarily make it possible to identify the nature of the activities and the personal involvement of the natural person in them*'.[288] It is settled case-law that '*the right to respect for private life with regard to the processing of personal data, recognised by Articles 7 and 8 of the Charter, concerns any information relating to an identified or identifiable individual.*'[289] It is therefore not the information disclosed *per se*, but its potential to identify the person concerned that makes it subject to the rights enshrined in Articles 7 and 8 of the Charter. Moreover, the Commission notes that, while the Union legislature explicitly recognised the potential impact on the rights to privacy and data protection of Article 39(1) of Regulation (EU) 2022/2065 by requiring providers to '*ensure that the repository does not contain any personal data of the recipients of the service to whom the advertisement was or could have been presented*', it did not consider it necessary to impose a similar obligation in relation to the personal data of advertisers.

(279) The provider of X does not explain in what way the information listed in Article 39(2)(c) of Regulation (EU) 2022/2065 in relation to third parties paying for advertisements [redacted] is sufficient to identify natural persons, the nature of

---

285 Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce'), OJ L 178, 17.7.2000, pages 1-16.

286 *See* Order of the Vice President of the Court of 27 March 2024, Amazon Services, C-639/23 P(R), EU:C:2024:277, paragraph 100.

287 Order of the President of the General Court of 2 July 2024, Aylo Freesites LTD v Commission, C-511/24, ECLI:EU:T:2024:431; Order of the President of the General Court of 12 July 2024, Commission v. WebGroup Czech Republic Case T-139/24 R, EU:T:2024:475.

288 *See* Order of the President of the General Court of 2 July 2024, Aylo Freesites LTD v Commission, C-511/24, ECLI:EU:T:2024:431, paragraph 112.

289 Judgment of 9 November 2010, Volker und Markus Schecke and Eifert, joined Cases C-92/09 and C 93/09, EU:C:2010:662, paragraph 52.

their activities, and the personal involvement of those natural persons in those activities. The Commission observes, in this regard, that Article 39(2)(b) and (c) of Regulation (EU) 2022/2065 only requires the disclosure of personal data in specific circumstances,[290] and the amount of personal data that the providers of very large online platforms and very large online search engines are required to disclose is limited (i.e., only the name of the natural or legal person).

(280) Any concerns that the provider of X may have as regards the potential to violate the privacy rights of third parties through disclosure of their names in the advertisement repository, insofar as it claims that it has 'no legal basis to do so', are in any event unfounded. Article 39 of Regulation (EU) 2022/2065 itself provides the legal basis for such disclosure. Any possible interference with the right to privacy of natural persons has been considered by the Union legislature in enacting that provision and are deemed necessary and proportionate to meet the objective of general interest pursued by that provision, namely to ensure research and supervision into risks stemming from online advertising.[291] The Commission considers that public disclosure of the information required by Article 39(2)(b) and (c) of Regulation (EU) 2022/2065 pursues a legitimate aim, as it holds advertisers accountable for their messaging and practices. The Commission also considers that, when advertisers know that information about their advertisements, including their legal name, will be accessible to the public, they are more likely to adhere to regulations and guidelines regarding advertising. Public scrutiny can also deter advertisers from engaging in misleading or deceptive advertising tactics. As regards the relationship between Article 39(2)(b) and (c) of Regulation (EU) 2022/2065 and the provider of X's obligations under Regulation (EU) 2016/679, the Commission observes that to the extent those data include personal data, Article 39(2) of Regulation (EU) 2022/2065 provides for a legal basis for such disclosure.

(281) In addition, the Commission notes that natural persons that advertise on X do so in a professional capacity.[292] A diligent trader should be aware of the disclosure obligations to which a provider of online platform are subjected under Regulation (EU) 2022/2065. Consequently, if a natural person acting as a trader takes the deliberate choice not to set up a legal entity through which to conduct its professional activities, it has accepted that its name will be disclosed when advertising on an online platform. The Commission further notes that the provider of X has not specified how the disclosure of the name of natural persons that advertise on X will negatively impact its business interests and relationships with advertisers, in particular under circumstances where a large number of very large online platforms and search engines that provide advertising services are subject to the same obligation.

(282) Fourth, as regards the provider of X's claim regarding the availability of commercial communications on X's advertisement repository, the Commission recalls that the technical limitations related to creation of the .csv file concerning such communications have been set out in detail in recital 204 above. As regards the provider of X's claim that the Commission failed to communicate to it any concerns it

290 Most notably in circumstances where the person on whose behalf the advertisement is presented, or, if applicable, the person who paid for the advertisement is a natural person.

291 Order of the Vice-President of the Court of 25 February 2025, WebGroup Czech Republic, a.s. v Commission, C-620/24, ECLI:EU:C:2025:136, paragraph 39.

292 *See* Order of the President of the General Court of 2 July 2024 Aylo Freesites LTD v Commission, C-511/24, ECLI:EU:T:2024:431.

had in relation to the Commercial Communications file prior to 6 November 2023, the Commission recalls that X's advertisement repository did not provide information on any commercial communications until at least 25 October 2023. Moreover, the Commission notes that legal obligations are applicable to the addressee irrespective of a notification to their addressee. Therefore, obligations addressed to the provider of X as the provider of a designated very large online platform under Regulation (EU) 2022/2065 are applicable regardless of whether the Commission or any third party has made the provider of X aware of their individual application. In sum, the provider of X must provide the Commercial Communications file regardless of whether it was aware of this obligation or not.

(283) Additionally, the provider of X, when publishing the Commercial Communications file on 6 November 2023, clearly failed to perform basic sanity checks on the file, such as trying to download, open and check it for duplicate entries, since it would have become immediately obvious to that provider that that file was too large to open fully with standard third-party software and that each entry was repeated 57 times in the file.

(284) Fifth, as regards the provider of X's claim that there are no serious or systemic issues compromising the reliability of the search tool, the Commission notes that it has documented evidence of a wide range of issues concerning the provider of X's search tool, related to a persistent and systemic malfunctioning.[293] As explained in recitals 199 to 203 above, the information output is consistently unreliable, which confirms that X's search tool persistently and systemically malfunctions. Moreover, as explained in recital 204 above, there was a considerable downtime period in October 2023 due to a typo, in the download link for the output. The provider of X has entirely failed to respond to these allegations in its Reply to the Preliminary Findings.

5.5.6. The claim that X's advertising repository is available through an API

(285) In the sixth place, as regards the provider of X's claim that X's advertisement repository can be accessed through an API and that it has not been made aware of any alleged dysfunctionalities with that API, there was in fact no effective access to X's API during the testing period as shown by the technical trials conducted by the Commission and further explained in recital 208 above.[294] In addition, the provider of X's claim that the Commission was unable to access the API because of a bug that was fixed on 13 July 2024 is incorrect, since during subsequent trials conducted by the Commission on 6 February 2025 to access the API demonstrated that it was still malfunctioning.[295] Moreover, during the trials described in recital 208 above, the Commission found that, contrary to what the provider of X claims, the API does refer to subscription tiers when accessed according to the instructions provided by that provider. This finding has been subsequently confirmed by third parties, as explained in recital 209 above.

---

293 X post, https://x.com/WhoTargetsMe/status/1714276498004435112, accessed on 19 June 2024 (DSA.100103, Doc ID 160-61). Correspondence between CheckFirst and X (DSA.100103, Doc ID 173-1). 'DSA.100103' Technical Analysis, section 2.2.2. (DSA.100103. Doc ID 219-1). Non-confidential minutes of interview with a representative from CheckFirst, 15 May 2024, (DSA.100103, Doc ID 166-1). Non-confidential minutes of interview with a representative from CheckFirst, 15 May 2024, (DSA.100103, Doc ID 166-1). See recital 205 above.

294 'DSA.100103' Technical Analysis (DSA.100103, Doc ID 219-1).

295 DSA.100103, Doc IDs 314, 315, 316, 317, 318 and 319.

(286) In addition, the provider of X's claim that its lack of awareness of the dysfunctional access to its API is factually incorrect. The Commission's decision to initiate proceedings on 18 December 2023 explicitly mentioned the failure to provide APIs to obtain all the information on advertisements presented on X's interface.[296] The provider of X acknowledged the receipt of this Decision on 22 December 2023. Similarly, as explained in recital 210 above, the provider of X is also wrong to claim that it was not made aware by third parties about the dysfunctionality of its API provision.

5.6. Conclusion

(287) For all the reasons set out in section 5.3 and 5.5. above, the Commission concludes that the provider of X has failed to comply with Article 39 of Regulation (EU) 2022/2065 by not providing a searchable and reliable tool that allows multicriteria queries of at least all the information in Article 39(2) of that Regulation, as well as by not providing an API to obtain all the information listed in Article 39(2) of Regulation (EU) 2022/2065 on such advertisements.

6. NON-COMPLIANCE WITH ARTICLE 40(12) OF REGULATION (EU) 2022/2065

6.1. The legal framework

(288) Pursuant to Article 40(12) of Regulation (EU) 2022/2065, '*[p]roviders of very large online platforms and of very large online search engines shall give access without undue delay to data, including, where technically possible, to real-time data, provided that the data is publicly accessible in their online interface by researchers, including those affiliated to not for profit bodies, organisations and associations, who comply with the conditions set out in paragraph 8, points (b), (c), (d) and (e) of that Article, and who use the data solely for performing research that contributes to the detection, identification and understanding of systemic risks in the Union pursuant to Article 34(1)*' of that Regulation.

(289) As specified in Article 40(12) of Regulation (EU) 2022/2065, providers of very large online platforms and of very large online search engines shall make access available to publicly accessible data to all researchers, including those researchers affiliated to not for profit bodies, organisations and associations, provided that those researchers use the data for the purposes described above and that they fulfil the eligibility conditions laid down in Article 40(8), points (b), (c), (d) and (e) of that Regulation, namely that:

> *'(b) they are independent from commercial interests;*
>
> *(c) their application discloses the funding of the research;*
>
> *(d) they are capable of fulfilling the specific data security and confidentiality requirements corresponding to each request and to protect personal data, and they describe in their request the appropriate technical and organisational measures that they have put in place to this end'; and*
>
> *(e) their application demonstrates that their access to the data and time frames are necessary for, and proportionate to, the purposes of their research, and that the expected results of that research will contribute to the purposes laid down in paragraph 4.*'

[296] Commission Decision C (2023) 9137 final, paragraph 14.

Researchers that fulfil such conditions shall hereinafter be referred to as 'Qualified Researchers'.

(290) Recital 98 of Regulation (EU) 2022/2065 further clarifies that, '*where data is publicly accessible, providers of very large online platforms or of very large online search engines should not prevent researchers meeting an appropriate subset of criteria from using this data for research purposes that contribute to the detection, identification and understanding of systemic risks. They should provide access to such researchers including, where technically possible, in real-time, to the publicly accessible data, for example on aggregated interactions with content from public pages, public groups, or public figures, including impression and engagement data such as the number of reactions, shares, comments from recipients of the service. Providers of very large online platforms or of very large online search engines should be encouraged to cooperate with researchers and provide broader access to data for monitoring societal concerns through voluntary efforts, including through commitments and procedures agreed under codes of conduct or crisis protocols. Those providers and researchers should pay particular attention to the protection of personal data, and ensure that any processing of personal data complies with Regulation (EU) 2016/679. Providers should anonymise or pseudonymise personal data except in those cases that would render impossible the research purpose pursued.'*

6.2. The relevant measures put in place by the provider of X

(291) During the months preceding the moment at which Regulation (EU) 2022/2065 started applying to X, i.e. 28 August 2023, the provider of X took decisions to dismantle Twitter's Academic API program, which was available to researchers from 2021 until the Spring of 2023, without providing any alternative.[297] Prior to these changes, the provider of Twitter had been successfully providing researchers access to data since 2006. [298]

(292) On 9 November 2023, the provider of X implemented a new application process for researchers seeking data access under Article 40(12) of Regulation (EU) 2022/2065. Hence, the assessment of the compliance by the provider of X with that provision can

---

297 See Nature, https://www.nature.com/articles/d41586-023-00460-z, accessed on 25 June 2024 (DSA.100102, Doc ID 207-42); Le Monde, https://www.lemonde.fr/sciences/article/2023/11/28/x-terrain-d-enquete-deserte-des-chercheurs_6202801_1650684.html, accessed on 25 June 2024 (DSA.100102, Doc ID 207-34); The Verge, https://www.theverge.com/2023/5/31/23739084/twitter-elon-musk-api-policy-chilling-academic-research, accessed on 19 June 2024 (DSA.100102, Doc ID 207-157); Fast Company, https://www.fastcompany.com/91040397/under-elon-musk-x-is-denying-api-access-to-academics-who-study-misinformation, accessed on 19 June 2024 (DSA.100102, Doc ID 207-105); Politico, https://www.politico.eu/article/elon-musk-twitter-goes-to-war-with-researchers-api/, accessed on 19 June 2024 (DSA.100102, Doc ID 207-93); CNN Business, https://edition.cnn.com/2023/04/05/tech/academic-researchers-blast-twitter-paywall/index.html, accessed on 19 June 2024, (DSA.100102, Doc ID 207-85).
One expert noted that X had well-functioning measures in place to provide researchers access to data prior to dismantling them in the spring of 2023.; See non-confidential minutes of interview with an expert, 10 January 2024 (DSA.100102, Doc ID 45-1).

298 Reply to the Preliminary Findings, footnote 286 (DSA.100102, Doc ID 337-2). This footnote mentions the EDMO report on researcher access to data; https://edmo.eu/wp-content/uploads/2022/02/Report-of-the-European-Digital-Media-Observatorys-Working-Group-on-Platform-to-Researcher-Data-Access-2022.pdf. This 2022 report contains a letter from Twitter which mentions that '*Since 2006, Twitter has had an open API allowing developers and researchers to use data from the public conversation to study diverse topics such as state-backed efforts to disrupt the public conversation, […]'* (DSA.100102, Doc ID 379).

be divided into two distinct time periods: the first from 28 August 2023 to 9 November 2023 and the second from 9 November 2023 onwards.

(293) For the period from 28 August 2023 to 9 November 2023, the provider of X claims to have put in place two solutions to provide Qualified Researchers with access to data pursuant to Article 40(12) of Regulation EU 2022/2065.[299] The first solution consisted in offering the possibility to subscribe to X's commercial '*developer API*' service.[300] The costs for each tier of that service ('*Free*', '*Basic*', '*Pro*', and '*Enterprise*') varied depending on the scope of the API access. This data access was not specifically offered by the provider of X to researchers for the purposes of complying with that provider's obligations under Article 40(12) of Regulation (EU) 2022/2065, but it was rather available to any commercial client who would be willing to pay for such access. The second solution consisted in the alleged possibility for researchers to request access from the provider of X via the email address EU-Questions@x.com, which was published on the EU-specific section of X's Help Center and presented as the general point of contact for all 'additional queries' that cannot be resolved only by consulting the Help Center.[301]

(294) For the period from 9 November 2023 onwards,[302] the provider of X – following the first RFI, in which the Commission noted that provider's apparent lack of willingness to give researchers access to publicly available data in accordance with Article 40(12) of Regulation (EU) 2022/2065 – implemented a new process which was meant to give access to eligible researchers through a dedicated application form for data access requests made pursuant to Article 40(12) of Regulation (EU) 2022/2065.[303] The provider of X claims that that application form (hereafter referred to as the '40(12) application form') was introduced '*to make it easier to intake applications from researchers while continuing to comply with the requirements of the DSA*'.[304] The provider of X claims that, upon approving a request for data access, it provides the researcher access to the X API to a level commensurate with the needs of the request.[305] The provider of X claims to provide approved researchers the '*Pro*' tier Developer API access level, the second-most extensive access scope,[306] free of

299 Reply to the second RFI, request 2 of section VII (DSA.100102, Doc ID 210-9).

300 The 'X Developer API' ('X API') can be used to programmatically retrieve and analyse X data, it is accessible to any individual or entity paying the costs set by the provider of X and accepting the relevant terms of service.; see https://developer.x.com/en, accessed on 21 June 2024 (DSA.100102, Doc ID 207-165).

301 This information was communicated to the Commission on 17 November 2023 in the reply to the first RFI, section 5 in parallel to the launch of the new application process discussed in recital 294 (DSA.100102, Doc ID 2-11).

302 The launch of the 40(12) application form coincided with the reply to the first RFI on 17 November 2023. The first RFI explicitly enquired about the specific dispositions of the provider of X to provide researchers access to data in accordance with Article 40(12) of Regulation (EU) 2022/2065 (DSA.100102, Doc ID 2-11).

303 Reply to the second RFI, request 2 of section VII (DSA.100102, Doc ID 210-9).

304 Reply to the second RFI, request 2 of section VII (DSA.100102, Doc ID 210-9).

305 Reply to the first RFI, section 5 (DSA.100102, Doc ID 2-11).

306 The Commission however notes that the superior 'Enterprise' tier, which starts at USD 42 000 per month, is a bespoke offering. This implies that an array of possible offerings allow for more extensive access than the 'Pro Tier'; See X Developer Platform, https://developer.x.com/en/products/twitter-api/enterprise/enterprise-api-interest-form, accessed on 19 June 2024 (DSA.100102, Doc ID 207-66).

charge.[307] In the first two months after introducing the form, the provider of X received 151 requests for access via the 40(12) application form.[308]

(295) The provider of X further claims that, from 9 November 2023 onwards, it has continued to improve its internal review process for requests made pursuant to Article 40(12) of Regulation (EU) 2022/2065 by: *'(1) preparing internal guidelines for the X employees in charge of reviewing the applications, (2) moving potentially eligible applications to a ticketing system for tracking correspondence, (3) working to understand and operationalize the Art. 40 criteria while waiting for further guidelines from the European Commission*'.[309]

(296) On [redacted] 2024, the first approval of researchers via the application process launched on 9 November 2023 took place, as [redacted] researchers saw their application being approved by the provider of X.

(297) As of 8 May 2024, the provider of X had received [redacted] requests through the process established on 9 November 2023. Out of those [redacted] requests, the provider of X had approved [redacted] requests, while it had rejected [redacted] requests, and [redacted] requests were still under review.[310]

(298) Over the entire relevant period from 9 November 2023 onwards, section 4 of X's terms of service stipulated the following : '*You may not do any of the following while accessing or using the Services: […] (iii) access or search or attempt to access or search the Services by any means (automated or otherwise) other than through our currently available, published interfaces that are provided by us (and only pursuant to the applicable terms and conditions), unless you have been specifically allowed to do so in a separate agreement with us (NOTE: crawling or scraping the Services in any form, for any purpose without our prior written consent is expressly prohibited)*'.[311] On 15 November 2023, the provider of X updated the X Developer Agreement, to include a reference to Article 40 of Regulation (EU) 2022/2065.[312] The X Developer

---

307 Reply to the second RFI, request 7 and 9 of section VII (DSA.100102, Doc ID 210-9).

308 This number corresponds to the period from 9 November 2023 to 9 January 2024 - See the table provided in reply to the fourth RFI, Request 32)b) (DSA.100102, Doc ID 160-5).

309 Reply to the second RFI, request 2 of section VII (DSA.100102, Doc ID 210-9).

310 See reply to the fourth RFI, request 32, including the relevant documents provided (DSA.100102; ID 160-2; 160-5; 168).

311 See X's terms of service: https://x.com/en/tos, accessed on 19 June 2024 (DSA.100102, Doc ID 207-46).

312 The X Developer Agreement are terms of service applying specifically to holders of '*X developer accounts*'. Accepting the X Developer Agreement is necessary to use the X API. Developer accounts are not necessary to use the X service, and as such, users are not required to accept the Developer Agreement to access the service. ; See Open Terms Archive, https://github.com/OpenTermsArchive/pga-versions/commit/c63b906e573f9f103649430a31a8080d2f7544e3#diff-2d91701cdd46e3d6db2a5cac7bd6531c806a0723a4f01ca8a1a6075eb98b533aR90, accessed on 19 June 2024 (DSA.100102, Doc ID 207-153). The added language reads: '***Digital Services Act****. Notwithstanding anything to the contrary in this Agreement, to the extent you are provided access to the Licensed Material pursuant to the procedures described in Article 40 of the Digital Services Act (The Digital Services Act) ("DSA"), your access and use of the Licensed Material is limited solely to performing research that contributes to the detection, identification and understanding of systemic risks in the European Union and only to the extent necessary for X to comply with its obligations under the DSA. Any such use of the Licensed Material is non-commercial as described in Section III(B) of this Agreement. You may not disclose, reproduce, license, or otherwise distribute the Licensed Material (including any derivatives thereof) that you retrieve through the X API to any person or entity outside the persons within your organization necessary to perform the research, unless (i) the information is*

Agreement solely applies to holders of X developer accounts, which are necessary to access the X API, in contrast to the X Terms of Service, which seemingly apply to anyone using X.

6.3. The Commission's Preliminary Findings

(299) In its Preliminary Findings, the Commission reached the preliminary conclusion that the provider of X, both before and after 9 November 2024, failed to comply with the obligation laid down in Article 40(12) of Regulation (EU) 2022/2065 to provide access to data that are publicly accessible on X's online interface to Qualified Researchers. This conclusion was based on the following considerations.

6.3.1. Period between 28 August 2023 and 9 November 2023

(300) As regards the period between 28 August 2023 and 9 November 2023, the provider of X claims to have put two solutions in place to provide Qualified Researchers with access to its publicly available data pursuant to Article 40(12) of Regulation EU 2022/2065. The first solution consisted in offering the possibility to subscribe to X's commercial 'developer API' service.[313] The second solution consisted in the possibility for researchers to request access via the email address EU-Questions@x.com.[314] In the Preliminary Findings, the Commission took the view that neither of those two solutions complied with Article 40(12) of Regulation (EU) 2022/2065.

(301) As regards the first solution, the Commission took the view that the practice of granting Qualified Researchers access to X's data through the purchase of X's commercial 'developer API' offering[315] deprived them of the ability to access data pursuant to Article 40(12) of Regulation (EU) 2022/2065.

(302) As reflected in Figure 10 below, the costs to access the 'developer API' amounted to USD 5 000 per month for the second most extensive 'Pro' access tier. While the provider of X also offered 'Free' and 'Basic' access tiers,[316] these lower tiers only allowed access to zero and 10 000 posts per month, respectively, which is a volume of data largely insufficient to provide a meaningful basis for most legitimate research projects conducted into systemic risks to which very large online platforms could give rise, considering that the average daily number of posts on X is estimated to be several hundred millions. Indeed, research related to large social media platforms regularly features large sample-sizes due to the significant quantities of content and underlying metrics (such as pertaining to engagement) associated with large social media

---

disclosed to the Digital Services Coordinator or other party specifically permitted by the DSA pursuant to the 'vetted researcher' status and procedures described in Article 40, or (ii) disclosure is required by law.'.

313 The 'X Developer API' ('X API') can be used to programmatically retrieve and analyse X data, it is accessible to any individual or entity paying the costs set by the provider of X and accepting the relevant terms of service.; see https://developer.x.com/en, accessed on 21 June 2024 (DSA.100102, Doc ID 207-165).

314 Reply to the second RFI, request 2 of section VII (DSA.100102, Doc ID 210-9).

315 The pricing of the 'X's Developer API' offering applies indiscriminately to for-profit business clients of the provider of X and public interest researchers, see Figure 10 below for an overview of the various API offerings. See also X Developer Platform, https://developer.x.com/en, accessed on 21 June 2024 (DSA.100102, Doc ID 207-165); X Developer Platform, https://developer.x.com/en/docs/twitter-api/getting-started/about-twitter-api, accessed on 19 June 2024 (DSA.100102, Doc ID 207-58).

316 An overview of the relevant access tiers is provided in Figure 10; The 'Free' access tier only permits uploads, and does not permit downloads.

platforms. [317] In this respect, the ability to access 10 000 post per month under the 'Basic' access tier is insufficient to enable any meaningful research associated with the study of social media platforms.[318]

(303) The Commission further observes that, for some research projects, even the scope of 'Pro' access tier is insufficient,[319] thus resulting in even higher costs associated with subscribing to the 'Enterprise' access tier, which starts at USD 42 000 per month.[320]

(304) The introduction of the aforementioned tier structure by the provider of X led to the documented cessation or cancellation of over a hundred research projects from April to November 2023.[321] Those documented projects represent only a fraction of all the research being performed on services from the provider of X. Cancelled or halted projects included long-term research projects which were recipients of Union funding (including European Research Council grants), and had been designed based on the ability to access X's publicly accessible data.[322] The Commission observes that the researchers who may have eventually regained access to the X API would not have been able to entirely make up for the time lost, since the data relevant for performing

---

317 See for example the following research papers, which rely on the analysis of hundreds of thousands, or and even millions of individual posts as well as other underlying metrics.
Barberá, Pablo, et al. 'Tweeting from left to right: Is online political communication more than an echo chamber?.' Psychological science 26.10 (2015): 1531-1542. (DSA.100102, Doc ID 256-1).
Vosoughi, Soroush, Deb Roy, and Sinan Aral. 'The spread of true and false news online.' science 359, no. 6380 (2018): 1146-1151. (DSA.100102, Doc ID 254-1).
Renault, Thomas, David Restrepo Amariles, and Aurore Troussel. 'Collaboratively adding context to social media posts reduces the sharing of false news.' arXiv preprint arXiv:2404.02803 (2024). (DSA.100102, Doc ID 255-1).

318 See non-confidential minutes of interview with researchers, 23 February 2024 (DSA.100102, Doc ID 78-1): '*Since the academic API was discontinued by X, the researchers have paid for the 'Basic' tier of the X API, which costs USD 100 per month and only enables them to 'scope' potential research projects, but not to conduct the research.*'.
See non-confidential minutes of interview with a researcher, 16 February 2024 (DSA.100102, Doc ID 103-1): '*Since then, he has been paying about a hundred euros a month for accessing the 'basic' tier of the API, which is only enough to conduct 'a skeleton of research', not more. He notes that this API tier does not allow him to gather data on a scale sufficient to conduct research, for which he would at minimum need access to the 'Pro tier' of the X API. However, he notes that paying for the 'Pro Tier' is simply not possible for him unless he obtained a large grant.*'.
See also testimony provided by a researcher, 26 February 2024 (DSA.100102, Doc ID 112-4): '*the official API is unaffordable, […] [redacted organisation name], like other civil society organizations and researchers which do not have a business model centred on selling data, simply cannot afford this.*'.

319 The suitability of the 'Pro Tier' for performing research on systemic risks is further examined in section 6.3.2.3.2. below.

320 The costs for 'Enterprise' API access are referenced in the following webpage: X Developer Platform, https://developer.x.com/en/products/twitter-api/enterprise/enterprise-api-interest-form, accessed on 19 June 2024 (DSA.100102, Doc ID 207-66).

321 This is a conservative estimate, based on the non-confidential results of a survey provided by a researcher, 26 February 2024 (DSA.100102, Doc ID 63-1).
See also Reuters, https://www.reuters.com/technology/elon-musks-x-restructuring-curtails-disinformation-research-spurs-legal-fears-2023-11-06/, 6 November 2023, accessed on 11 June 2024 (DSA.100102, Doc ID 184-1).

322 For concrete examples of how the discontinuation of this program by the provider of X affected researchers, see the non-confidential minutes of interviews with researchers on 19 March 2024, 1 March 2024, 23 February 2024, 26 February 2024, 22 January 2024, 8 February 2024 (DSA.100102, IDs 174-1; 116-1; 78-1; 86-1; 36-1; 133-1).

research on systemic risks is regularly deleted within weeks of publication (e.g. data related to disinformation, hate speech, inauthentic accounts).[323]

(305) In addition, a large number of researchers have publicly and privately declared that they were unable to sustain the financial costs associated with accessing X's 'developer API' product in a manner compatible with performing research and had to halt or cancel ongoing and upcoming research projects.[324] Finally, the Commission notes that the providers of other very large online platforms whose core features involve publicly accessible data give Qualified Researchers free of charge access to existing commercial APIs or to new APIs rolled out by those providers specifically for Qualified Researchers.[325]

Figure 10: comparing the access tiers available under the commercial offering of the 'X Developer API' Note: The Enterprise tier starts at USD 42 000 per year.

	Free	Basic	Pro	Enterprise
Getting access	Get Started	Get Started	Get Started	Get Started
Price	Free	$100/month	$5000/month	
Access to X API v2	✔ (Only Post creation)	✔	✔	
Access to standard v1.1	✔ (Only Media Upload, Help, Rate Limit, and Login with X)	✔ (Only Media Upload, Help, Rate Limit, and Login with X)	✔ (Only Media Upload, Help, Rate Limit, and Login with X)	
Project limits	1 Project	1 Project	1 Project	
App limits	1 App per Project	2 Apps per Project	3 Apps per Project	
Post caps - Post	1,500	3,000	300,000	
Post caps - Pull	✘	10,000	1,000,000	
Filtered stream API	✘	✘	✔	
Access to full-archive search	✘	✘	✔	
Access to Ads API	✔	✔	✔	

Source: X Developer Platform, https://developer.x.com/en/docs/twitter-api/getting-started/about-twitter-api, accessed on 19 June 2024 (DSA.100102, Doc ID 207-58).

(306) For these reasons, the Commission took the preliminary view that, between 28 August 2023 and 9 November 2023, the costs associated with obtaining access to publicly available data, in particular in the absence of any dedicated alternative for accessing such data, suppressed the ability of Qualified Researchers to access and use data that is

323 Non-confidential minutes of interviews with researchers, 1 March 2024 (DSA.100102, Doc ID 116-1).

324 Non-confidential minutes of interviews with researchers, 1 March 2024, 26 February 2024, 22 January 2024, 23 February 2024, 22 January 2024, 16 February 2024 (DSA.100102, Doc ID 116-1; 86-1; 36-1; 78-1; 40-1; 103-1).
Non-confidential testimony of a researcher, 27 February 2024 (DSA.100102, Doc ID 91-1).

325 EU-US Trade and Technology Council, Status Report: Mechanisms for Researcher Access to Online Platform Data, 5 April 2024, pp 12-17 (DSA.100102, Doc ID 185-1); See for instance the cases of YouTube, Facebook, Instagram, TikTok, LinkedIn.

publicly accessible in X's online interface to conduct research contributing to the detection, identification and understanding of the systemic risks in the Union, in clear contravention of Article 40(12) of Regulation (EU) 2022/2065.

(307) As regards the second solution the provider of X claims to have put in place to ensure access to data from 28 August 2023 to 9 November 2023, the Commission took the preliminary view that the possibility for researchers to request access to data via the email EU-Questions@x.com,[326] was ineffective, since it was not communicated by the provider of X to the potential beneficiaries of that solution. In the absence of any public communication, researchers were effectively deprived of the ability to benefit from their right to data access under Article 40(12) of Regulation (EU) 2022/2065.

(308) The absence of a public communication from the provider of X contrasts with the extensive media coverage on the discontinuation of the Twitter Academic API program, which drew wide public criticism from the research community.[327] Before November 2023, the provider of X made public statements that pointed all prospective users of the X API (including researchers) to access data through the commercial developer API subscription and did not provide any subsequent information regarding the introduction of a research-specific access to data, instead of issuing any form of clear communication about the possibility to apply for access pursuant to Article 40(12) of Regulation (EU) 2022/2065.[328] This is also illustrated by the fact that none of the researchers interviewed by the Commission was aware of any way to apply for

---

326 This information was communicated to the Commission on 17 November in the reply to the first RFI, section 5 in parallel to the launch of the new application process on 9 November 2023 (DSA.100102, Doc ID 2-11).

327 See Nature, https://www.nature.com/articles/d41586-023-00460-z, accessed on 25 June 2024 (DSA.100102, Doc ID 207-42).
Le Monde, https://www.lemonde.fr/sciences/article/2023/11/28/x-terrain-d-enquete-deserte-des-chercheurs_6202801_1650684.html, accessed on 25 June 2024 (DSA.100102, Doc ID 207-34).
The Verge, https://www.theverge.com/2023/5/31/23739084/twitter-elon-musk-api-policy-chilling-academic-research, accessed on 19 June 2024 (DSA.100102, Doc ID 207-157).
Fast Company, https://www.fastcompany.com/91040397/under-elon-musk-x-is-denying-api-access-to-academics-who-study-misinformation, accessed on 19 June 2024 (DSA.100102, Doc ID 207-105).
Politico, https://www.politico.eu/article/elon-musk-twitter-goes-to-war-with-researchers-api/, accessed on 19 June 2024 (DSA.100102, Doc ID 207-93).
CNN Business, https://edition.cnn.com/2023/04/05/tech/academic-researchers-blast-twitter-paywall/index.html, accessed on 19 June 2024, (DSA.100102, Doc ID 207-85).

328 X announced its revamped commercial API offering on 30 March 2023. At the time, the provider of X did not announce any researcher access channel to replace the 'Academic API' access. To the knowledge of the Commission, the provider of X did not make any additional communications to researchers until November of 2023, when it updated one of its webpages.
In its announcement of the new API offering on 30 March 2023, the provider of X mentioned that '*For Academia, we are looking at new ways to continue serving this community. In the meantime Free, Basic and Enterprise tiers are available for academics. Stay tuned to @TwitterDev to learn more*.'; See X post from the official 'X Developer', https://x.com/XDevelopers/status/1641222788911624192, accessed on 19 June 2024 (DSA.100102, Doc ID 207-74).
See also this blog, which analyses the new offering launched in the Spring of 2023 and mentions that '*There are no special access plans for researchers and academics at this time*.'; Superface AI, https://superface.ai/blog/twitter-api-new-plans, accessed on 19 June 2024 (DSA.100102, Doc ID 207-109).
See examples of researchers seeking information since the Spring of 2023 in the 'academic research' tab of X's online developer forums. Developers working for Twitter/X regularly replied to researchers on these forums under the Twitter Academic API program but have stopped doing so between the spring of 2023 and the Preliminary Findings; X developers, https://devcommunity.x.com/c/academic-research/62, accessed on 29 May 2024 (DSA.100102, Doc ID 181-1; 182-1; 183-1).

free API access for research purposes until the provider of X made an online form available on 9 November 2023.[329] This is further corroborated by media coverage of the introduction of the 40(12) application form on 9 November 2023, which was widely perceived as a decision by the provider of X to resume the intake of researcher applications in November 2023 for the first time since discontinuing the Twitter Academic API program.[330]

(309) The Commission therefore took the preliminary view that, between 28 August 2023 and 9 November 2023, the alleged possibility to apply via EU-Questions@x.com was never effectively communicated by X to researchers, thereby depriving researchers of the chance to benefit from it and rendering it ineffective. Over the relevant period, the only effective possibility for Qualified Researchers to access X's data was therefore to purchase access to X's 'developer API' service which, however, for the reasons explained above, did not meet the requirements of Article 40(12) of Regulation (EU) 2022/2065.

(310) Finally, as regards both solutions, the Commission took the view that the provider of X did not have any defined process in place for assessing applications pursuant to Article 40(12) of Regulation (EU) 2022/2065 between 28 August 2023 and 9 November 2023. The very first version of the internal workflow provided by the provider of X regarding the assessment of researcher applications pursuant to Article 40(12) of Regulation (EU) 2022/2065 dates from November 2023.[331] Furthermore, the version of the processes dating back to November 2023 appears to be only partially drafted, as it was for instance lacking any clear process to approve applications and onboard applicants, signalling the company's unpreparedness to do so.[332] A more definitive and comprehensive version of this workflow was finalized only in January 2024.[333] This observation is further corroborated by the fact that none of the 143 applications submitted by researchers between 9 November 2023 and 31 December 2023 were approved by the provider of X before 26 January 2024.

(311) For all the foregoing reasons, the Commission took the preliminary view that the provider of X failed to comply with Article 40(12) of Regulation (EU) 2022/2065 during the period 28 August 2023 to 9 November 2023.

6.3.2. Period following 9 November 2023

(312) As regards the period after 9 November 2023, the Commission took the preliminary view that the changes implemented by the provider of X to receive and assess

---

329 The interviews of researchers conducted by the Commission show that until the possibility to apply via the 40(12) application form became known only from late November 2023, researchers were not aware of any possibility to access the X API as researchers other than paying for the commercial access tiers; See for instance the non-confidential minutes of interviews with researchers on 19 March 2024, 1 March 2024, 23 February 2024, 26 February 2024, 22 January 2024, 16 February 2024, 8 February 2024, 28 February 2024 (DSA.100102, Doc ID 174-1; 116-1; 78-1; 86-1; 36-1; 103-1; 133-1; 113-1). See also the non-confidential version of the testimony provided by a researcher, 26 February 2024 (DSA.100102, Doc ID 112-4).

330 See for example TechCrunch, https://techcrunch.com/2023/11/17/change-in-xs-terms-indicate-eu-researchers-will-get-api-access/, accessed on 19 June 2024. (DSA.100102, Doc ID 207-89); X Post by Mathias Vermeulen, https://x.com/mathver/status/1729894799954600411, accessed on 19 June 2024 (DSA.100102, Doc ID 207-117).

331 See the templates provided in reply to the second RFI, request 5, section VII (DSA.100102, Doc ID 210-9; 210-7; 210-8).

332 Annexes provided in reply to the second RFI, Request 5 of section VII (DSA.100102, Doc ID 210-9).

333 See also section 6.3.2.2.1 below.

researcher applications, and to subsequently provide access to Qualified Researchers, did not bring the provider of X in compliance with the obligation laid down in Article 40(12) of Regulation (EU) 2022/2065.

(313) More specifically, the Commission noted significant shortcomings in the processes put in place by the provider of X to receive and assess researchers' 40(12) applications as of 9 November 2023, as well as to give Qualified Researchers adequate access to data via the X API in a manner which would allow performing research contributing to the detection, identification and understanding of the systemic risks stemming from the design and functioning of X in the Union. As explained below, these shortcomings exist at all stages of the process implemented by the provider of X to give Qualified Researchers access to data: (i) prior to the application; (ii) at the submission stage; (iii) at the assessment stage; (iv) at the outcome stage; and (v) in relation to providing adequate access to successful applicants.

6.3.2.1. Overly restrictive interpretation of the eligibility requirements set out in Article 40(12) of Regulation (EU) 2022/2065

(314) In the first place, the Commission took the preliminary view that the provider of X interpreted the eligibility requirements set out in Article 40(12) of Regulation (EU) 2022/2065 in an overly restrictive manner, thus restricting access for some of the eligible beneficiaries pursuant to that Article.

6.3.2.1.1. Eligible use of publicly accessible data

(315) First, the provider of X has an overly restrictive interpretation of the requirement that researchers use the data 'solely' for performing research that contributes to the detection, identification and understanding of systemic risks in the Union within the meaning of Article 34(1) of Regulation (EU) 2022/2065.

(316) As of 8 May 2024, [redacted] out of [redacted] applications for data access rejected by the provider of X were rejected on the basis that '*it does not appear that* [the researchers'] *proposed use of X data is solely for performing research that contributes to the detection, identification and understanding of systemic risks in the EU as described by Article 34 of* [Regulation (EU) 2022/2065]'.[334]

(317) This restrictive interpretation is corroborated by evidence collected by researchers affiliated with the European New School and the Weizenbaum Institute in their ongoing survey of the research community in relation to platforms' data access mechanisms.[335] The survey report identifies the repeated use of this argument by the provider of X to reject applications and notes that '*while this wording is found in Article 40(12), it is an unnecessarily narrow interpretation of the provision to restrict data access by claiming a request is not specific enough without further explanations*'.

(318) When deciding to escalate an application (i.e., asking the applicant for additional information),[336] which is part of the provider of X's mandatory approval track and

334 See the table provided in reply to the fourth RFI, Request 32)b) (DSA.100102, Doc ID 160-5); This search can be reproduced by filtering Column R to only include reasons which include 'Research outside scope'.

335 Weizenbaum Institute, https://www.weizenbaum-library.de/items/39a490c3-f3c1-42a7-a530-318b17e9de49, April 2024, in particular 'Sample Case 1' on Page 5, and page 6 of the report (DSA.100102, Doc ID 128-1).

336 Annex provided in reply to the second RFI, Request 5 of section VII; Pages 9 and 10 of the 'Version 2.0' Checklist (DSA.100102, Doc ID 210-9; 210-7; 210-8).

happens only if '*the initial information provided in the form seems legitimate and to warrant API access*', the guidance for X staff in charge of assessing whether the scope of research fits the criteria recommends them to '*ensure we have enough information about the research to confirm that the data will be used solely* [underlined in the original internal document] *for the purposes permitted in Articles 34 and 40*'. The emphasis placed on 'solely' in this document suggests that the internal instruction for X employees is to be particularly strict on this specific requirement when assessing requests.

(319) To support its Preliminary Findings, the Commission provided a non-exhaustive list of 12 specific examples of applications that were ultimately rejected by the provider of X on the unique ground that the applicant's proposed use of X data was allegedly not solely for '*performing research that contributes to the detection, identification and understanding of systemic risks in the EU as described by Article 34'*, which in fact appeared to fulfil this requirement.[337] The Commission conducted an in-depth analysis of 4 of these 12 rejections,[338] including by scrutinising the internal documents of the provider of X containing the reasoning (or lack thereof) of its technical staff to reject these 4 applications, which had never been fully communicated to the rejected applicants. [339]

(320) In response to the Commission's enquiry into the relevant qualifications of employees of the provider of X in charge of assessing applications, the provider of X mentioned only three employees working part - time on assessing data access requests.[340] These employees were claimed to possess [redacted], [redacted] and [redacted] years of 'relevant work experience', yet no indication was provided of their familiarity with the state of the art of research on systemic risks in the Union, which contrasts with the expertise of the researchers whose applications they were in charge of assessing. This mismatch of expertise is corroborated in the Commission's Technical Analysis of 4 specific applications supporting the Preliminary Findings on DSA.100102 ('the Technical Analysis'), which points to important flaws in the internal arguments[341] of the provider of X's staff to deny applications based on their scope. In addition, a majority of researchers who have been rejected by the provider X on this basis and who were in contact with the Commission disagreed with that provider's claims that the research project described in their applications did not fulfil this specific requirement.

---

337 Namely Applicants [redacted]; See the 'DSA.100102' Technical Analysis which provides an analysis of specific rejections, including in relation to research projects which were recipients of EU funding; 'DSA.100102' Technical Analysis (DSA.100102, Doc ID 257-1) and the table provided in reply to the fourth RFI, Request 32)b) (DSA.100102, Doc ID 160-5).

338 'DSA.100102' Technical Analysis (DSA.100102, Doc ID 257-1).

339 The shortcomings related to the lack of effective substantiation of rejections by the provider of X are further discussed in section 6.3.2.2.2.

340 Based on the reply to the second RFI, request 6)f of section VII: '*X currently has three employees involved in the day-to-day review of applications and in the ensuing correspondence with researchers'. The time devoted by these employees to the review of applications 'varies day-to-day', as these employees perform other duties in parallel within TIUC. These three employees consist in a* [redacted] *as well as a* [redacted] *and an* [redacted] *supporting the X Developer API*' (DSA.100102, Doc ID 210-9).

341 These arguments were internal to the provider of X and were never disclosed to rejected applicants. They are discussed extensively in section 6.5.3.1 of this document. See also 'DSA.100102' Technical Analysis (DSA.100102, Doc ID 257-1).

(321) The Commission observes that the notion '*solely*' to which Article 40(12) of Regulation (EU) 2022/2065 refers should not be understood as referring to the scope of the research project or scope of data to which access is to be granted, but to the purpose of such access, i.e., '*performing research that contributes to the detection, identification and understanding of systemic risks in the Union pursuant to Article 34(1)'* of Regulation (EU) 2022/2065. Therefore, the wording of that provision does not require that the scope of the specific research project in question be limited to the detection, identification and understanding of systematic risks in the Union, so long as the final use of the data in question is solely for the purpose of performing research that contributes to the detection, identification and understanding of systemic risks in the Union pursuant to Article 34(1) of Regulation (EU) 2022/2065. The Commission takes the view that this interpretation is further supported by the requirement that the research 'contributes' to this purpose, which can be accomplished by a broad set of research projects and by a broad use of data samples.

(322) Even when it does not lead to a rejection of the applicants, the restrictive interpretation of Article 40(12) of Regulation (EU) 2022/2065 adopted by the provider of X artificially restricted the scope of the research projects. In fact, in this respect, one of the [redacted] successful applicants as of 8 May 2024 explained to the Commission that they were cognizant of the provider of X's very restrictive interpretation regarding acceptable data use. Considering this, they claim to have built their data access request in a very '*conservative*' manner that would be '*impossible for X to reject*'.[342] The researcher did so by formulating a proposal to study the prevalence and the amplification of misinformation in the context of [redacted]. During the application review process, the provider of X requested, on three consecutive occasions, that researcher to reaffirm and provide evidence that the applicant's proposed research would indeed focus on '*misinformation in the context of the elections as opposed to misinformation more broadly*'. This exemplifies that the restrictive interpretation by the provider of X also had an impact on the formulation of research proposals, namely by incentivising researchers to limit the scope of their projects further than initially planned and beyond what would be strictly needed to be granted access to data by the provider of X, had the latter properly applied Article 40(12) of Regulation (EU) 2022/2065.

(323) For all these reasons, the Commission took the preliminary view that the provider of X reviewed and rejected certain applications based on an overly restrictive interpretation of the requirement that data be used *'solely for performing research that contributes to the detection, identification and understanding of systemic risks in the Union pursuant to Article 34(1)'.* By doing so, the provider of X has therefore failed to comply with Article 40(12) of Regulation (EU) 2022/2065.

6.3.2.1.2. Restrictions based on the geographic location of the researchers

(324) Second, the provider of X rejected researchers on the sole basis that they were established outside the Union, whereas Article 40(12) of Regulation (EU) 2022/2065 does not allow for such a restriction.

(325) The Commission is specifically aware of two such instances in which the provider of X rejected data access requests coming from researchers established outside the Union on the sole basis of the establishment of those researchers (using language such as '*X*

[342] See the non-confidential minutes of interviews with a researcher on 28 February 2024 (DSA.100102, Doc ID 113-1).

will limit API access under Art. 40(12) to organizations located in the EU. Accordingly, your application has been denied.'), despite the fact that the subject-matter of these requests was concerned with performing research that contributes to the detection, identification and understanding of systemic risks in the Union. Moreover, the provider of X rejected those applications after having asked those researchers for extensive additional information supporting their requests, despite being aware that those researchers were affiliated to non-EU universities since the very first moment that they applied, and having failed to reply to both applicants for more than a month after they had provided the additional information requested by the provider.[343]

6.3.2.1.3. (Non-)Affiliation to a research organisation

(326) Third, the application form implemented by the provider of X features two mandatory fields related to the researchers' organisation[344] and internal documents submitted to the Commission by the provider of X instructs employees to deny applications when applicants are not affiliated to a '[redacted]'.[345] However, Article 40(12) of Regulation (EU) 2022/2065 does not require Qualified Researchers to be affiliated to an organisation, body or association, in contrast to Article 40(4) of that Regulation where the Union legislature explicitly required researchers to be affiliated to a research organisation to obtain the status of 'vetted researcher' and submit reasoned requests to access data under Article 40(8) of that Regulation. The mandatory nature of those fields in the application form, and the internal guidelines to reject applicants on the basis of their affiliation thus entail a restriction that goes further than the scope of beneficiaries defined in Article 40(12) of Regulation (EU) 2022/2065.[346]

6.3.2.1.4. Disclosure of funding and independence from commercial interests

(327) Fourth, the Commission took the view that the application form of the provider of X requires the provision of information going beyond what is needed to verify the subset of criteria laid down in Article 40(12) of Regulation (EU) 2022/2065. This often leads the provider of X to request information from researchers that is not required to assess whether the researcher fulfils the conditions for data access under Article 40(12) of Regulation (EU) 2022/2065.

---

343 See in table provided in reply to the fourth RFI, request 32)b) (DSA.100102, Doc ID 160-2; 160-5).
Applicant [redacted] provided additional information on [redacted] 2024 and did not receive a reply from the provider of X until [redacted] 2024, despite asking for updates. His request was denied on the basis that '*until we receive any contrary legislation or guidance, X will limit API access under Art. 40(12) to organizations located in the EU. Accordingly, your application has been denied*.'. The applicant complained that this was not stated earlier and that the provider of X made him reply to several waves of questions over the course of four months. He insisted that his request focused on systemic risks to the European Union.
Applicant [redacted] provided additional information on [redacted] 2024 and did not receive a reply until [redacted] 2024. His request was also rejected on the sole basis that '*Based on your application, it appears your organization is outside of the European Union.*', after the applicant had been asked additional information by X, and as is proposed research focused on assessing the prevalence of hate speech in the EU.

344 Namely, the '*Name of Organization*' and '*Organization Affiliation*' fields.

345 'V2.2 DSA Art 40(12) Research access checklist', page 3, provided in reply to the fourth RFI, request 33 (DSA.100102, Doc ID 160-2; 168-604).

346 To be granted the vetted status pursuant to Article 40(4) of the Digital Services Act, researchers have to be affiliated to a research organisation as defined in Article 2, point (1), of Directive (EU) 2019/790.

(328) The provider of X claims that its '*40(12) Application form tracks the requirements*' listed in Article 40(12) of Regulation (EU) 2022/2065,[347] namely that the researchers fulfil '*the criteria in Article 40(8), points (b), (c), (d) and (e), and that they will use the data solely for performing research that contributes to the detection, identification and understanding of systemic risks in the Union pursuant to Article 34(1)*'. However, the provider of X has interpreted the conditions in Article 40(8) point (b) ('independence from commercial interests') and point (c) ('disclosing the funding of the research') to require the provision of extensive information related to the researcher's organisation and its board, [348] as well as information going beyond the funding of the research. [349]

(329) While the affiliation of an applicant to a specific organisation may be relevant to assess whether that applicant is entitled to request access to data pursuant to Article 40(12) of Regulation (EU) 2022/2065, there is nothing in that provision which would justify requesting mandatory information about the organisation's board members, membership, shareholders, grant recipients, or information about the 'indirect funding'[350] of researchers. Requesting that information is therefore designed to discourage meaningful scrutiny of the systemic risks to which a service gives rise. In a public blog post, Brandon Silverman, former CEO and co-founder of CrowdTangle,[351] who is a leading expert in providing researchers access to platform data, noted that these questions about '*indirect funding sources […] and detailed information about an organization's board members and directors*' go '*beyond traditional industry norms for researcher vetting for public data on social media*'. [352] In parallel, two researchers interviewed by the Commission shared their concerns that the provider of X was using this application form to map who is funding the research field as a whole. [353] This is particularly relevant in light of recent lawsuits in which the provider of X sued several critical civil society researchers. [354] Internal documents provided by the provider of X

---

347 Reply to the second RFI, request 4 of Section VII (DSA.100102, Doc ID 210-9).

348 The "Commercial Interests" box of the 40(12) application form requests applicants to disclose information, "including but not limited to […] (i) purposes and mission (ii) information about your board members, officers, and individuals with access to X data, (iii) previously published research, (iv) affiliation with other organizations, networks, or causes, and (v) the organization's membership, shareholders or grant recipients"; See https://docs.google.com/forms/d/e/1FAIpQLSdo0O-D6Kxa3cV4g1JLz2T_0Sk3hdEnTdv8dJmibagCnzJ7kg/viewform, accessed on 19 June 2024 (DSA.100102, Doc ID 207-70).

349 The 'Funding' box of the '40(12) Application form' requests applicants to disclose the 'funding of your research, including amounts raised and direct and indirect sources of funding.', https://docs.google.com/forms/d/e/1FAIpQLSdo0O-D6Kxa3cV4g1JLz2T_0Sk3hdEnTdv8dJmibagCnzJ7kg/viewform, accessed on 19 June 2024 (DSA.100102, Doc ID 207-70).

350 The '40(12) application form' implemented by the provider of X requires that applicants provide information on '*direct and indirect sources of funding*' (emphasis added) where 'indirect funding' appears to be ill defined and goes beyond the requirement set out by Article 40(8)(c) (namely, that the "application discloses the funding of the research"). ; See https://docs.google.com/forms/d/e/1FAIpQLSdo0O-D6Kxa3cV4g1JLz2T_0Sk3hdEnTdv8dJmibagCnzJ7kg/viewform, accessed on 19 June 2024 (DSA.100102, Doc ID 207-70).

351 CrowdTangle was a social monitoring tool, eventually acquired by Meta, that was used by thousands of civil society groups around the world to monitor public content on social media.

352 Mozilla Foundation, https://foundation.mozilla.org/en/blog/EU-Digital-Services-Act-and-The-State-of-X-Transparency/, accessed on 19 June 2024 (DSA.100102, Doc ID 207-161).

353 Non-confidential minutes of interview with researchers, 22 January 2024 (DSA.100102, Doc ID 40-1).

354 Washington Post, https://www.washingtonpost.com/technology/2024/03/25/musk-x-lawsuit-slapp-center-digital-hate/, accessed on 25 June 2024 (DSA.100102, Doc ID 207-38).

at the request of the Commission [355] contain language about the need to particularly scrutinise the commercial interests of non-academic organizations, '*because many of these are subject to commercial interests*'. However, the Commission notes that the form implemented by the provider of X for the purposes of researchers' applications under Article 40(12) of Regulation (EU) 2022/2065 contains these questions regardless of whether those applications originate from academic or non-academic researchers, which further indicates that some of the information requested in this form are disproportionate.

6.3.2.2. Shortcomings in the application review processes and in communicating outcomes to applicants

(330) In the second place, the Commission took the preliminary view that the processes set up by the provider of X to review applications to access publicly accessible data via the X API pursuant to Article 40(12) of Regulation (EU) 2022/2065, as well as to communicate outcomes to applicants, presented significant shortcomings.

6.3.2.2.1. Failure to give access to data without undue delay

(331) First, the Commission took the view that the provider of X repeatedly failed to assess applications and to give applicants access to data via the X API *without undue delay*, as required by Article 40(12) of Regulation (EU) 2022/2065.

(332) As explained in recital 310 above, the internal review procedures of the provider of X were only partially finalised in January 2024 at the earliest, which was more than four months after Regulation (EU) 2022/2065 started to apply to X. This contributed to significantly delaying the ability of Qualified Researchers to access data via the X API, and therefore their ability to perform research on systemic risks.

(333) According to internal documents, 'Version 1.0'[356] of the procedure to assess applications and provide access to successful applicants (used from November 2023 to January 2024) lacked a clear and defined process to approve applications and onboard applicants, which only appears in Version 2.0 (implemented from January 2024 onwards), signalling an unpreparedness to do so. This is consistent with the fact that, despite the launch of a new process to receive applications on 9 November by the provider of X, not a single application was approved before 26 January 2024, almost 5 months after Regulation (EU) 2022/2065 became fully applicable to the provider.[357]

(334) In addition, the procedure put into place by the provider of X to review applications regularly leads to unnecessary delays and therefore fails to give access to applicants *without undue delay* as required by Article 40(12) of Regulation (EU) 2022/2065.

(335) More specifically, the provider of X repeatedly failed to reply to applicants for more than a month and sometimes failed to reply until applicants enquired about the absence of updates.[358] In addition, the [redacted] researchers to whom the provider of X granted

355 Annexes provided in reply to the second RFI, Request 5 of Section VII (DSA.100102, Doc ID 210-9).

356 Annexes provided in reply to the second RFI, Request 5 of section VII (DSA.100102, Doc ID 210-9).

357 Table provided in reply to the fourth RFI, Request 32)b) (DSA.100102, Doc ID 160-2; 160-5).

358 See for instance applications [redacted] (in particular document TIUC-1331); [redacted] (in particular document TIUC-242 and -465); [redacted] (in particular document TIUC-1014); 24-42 (in particular document TIUC-1145); [redacted] (in particular document TIUC-367); [redacted]; [redacted]; [redacted]; [redacted] in the documents provided in reply to the fourth RFI, request 32 (DSA.100102, Doc ID 160-5; 168). See JIRA tickets provided in reply to the fourth RFI, Request 34, in particular: document TIUC-1739 – comment written on 3 January 2024 'Applicant [redacted] will receive an email on [redacted] 2024 informing them that their application is under review.' (DSA.100102, Doc ID 160-5; 168).

API access (the [redacted] successful applicant received data via a file transfer) were all informed of their approval on the same two dates,[359] pointing to inefficiencies in the workflow implemented by the provider of X for the application review process.

6.3.2.2.2. Explanations to rejected applicants do not properly substantiate the alleged shortcomings

(336) Second, the explanations provided by the provider of X to rejected applicants lacked detail, thus preventing applicants from understanding the reasons leading to their rejection, which has the effect of hindering their ability to challenge their rejection or to adapt their application with a view of obtaining access pursuant to Article 40(12) of Regulation (EU) 2022/2065 in the future.

(337) In the large majority of rejection emails sent by the provider of X to researchers, that provider adopted generic language pointing to an alleged incompleteness of the request or a failure to fulfil one of the applicable requirements.[360] Figure 11 below, demonstrates that the reply templates used by the provider of X[361] by default did not include space for any additional or specific feedback to the rejected researcher except for referencing alleged shortcomings with regard to one of the eligibility criteria laid down in Article 40(12) of Regulation (EU) 2022/2065. In fact, only [redacted] out of [redacted] rejected applicants received a 'custom' rejection message deviating from the generic wording contained in that template.[362]

(338) The emails that the provider of X sent to researchers when rejecting their applications included a note explaining that '*responses to this email are not monitored*', whereas in previous steps of the application process, researchers interacted with the provider of X by replying to emails using the same address as the one used to ultimately send the rejection ([redacted]@x.com).

(339) Recital 98 of Regulation (EU) 2022/2065 clarifies that providers '*should not prevent*' researchers meeting an appropriate subset of criteria from using publicly accessible data for research purposes that contribute to the detection, identification and understanding of systemic risks. The provider of X's failure to disclose to researchers meaningful indications as to the reasons for which their requests have been rejected is likely to deter such researchers from making use of their right.[363] In this respect,

---

359 Out of the [redacted] successful applicants who obtained access to the X API as of 8 May 2024, [redacted] decisions to grant access were communicated to applicants on [redacted] 2024 and [redacted] were communicated on [redacted] 2024. The [redacted] researcher who has been granted access saw their request approved on [redacted] 2024, their case differs from the [redacted] others as [redacted]. See the table provided in reply the fourth RFI, Request 32)b) (DSA.100102, Doc ID 160-2; 160-5).

360 As per the table provided in reply to the fourth RFI, request 32)b) only [redacted] out of [redacted] rejected applicants received a 'custom' denial (this can be replicated by filtering column R) (DSA.100102, Doc ID 160-2; 160-5). See the specific examples provided in the 'DSA.100102' Technical Analysis (DSA.100102, Doc ID 257-1).

361 See 'Version 2.0' of the research access checklist provided in Annex to the reply to the second RFI, Request 5 of section VII ; page 14 (DSA.100102, Doc ID 210-8). See also the reply to the second RFI, request 6)e) of section VII (DSA.100102, Doc ID 210-9).

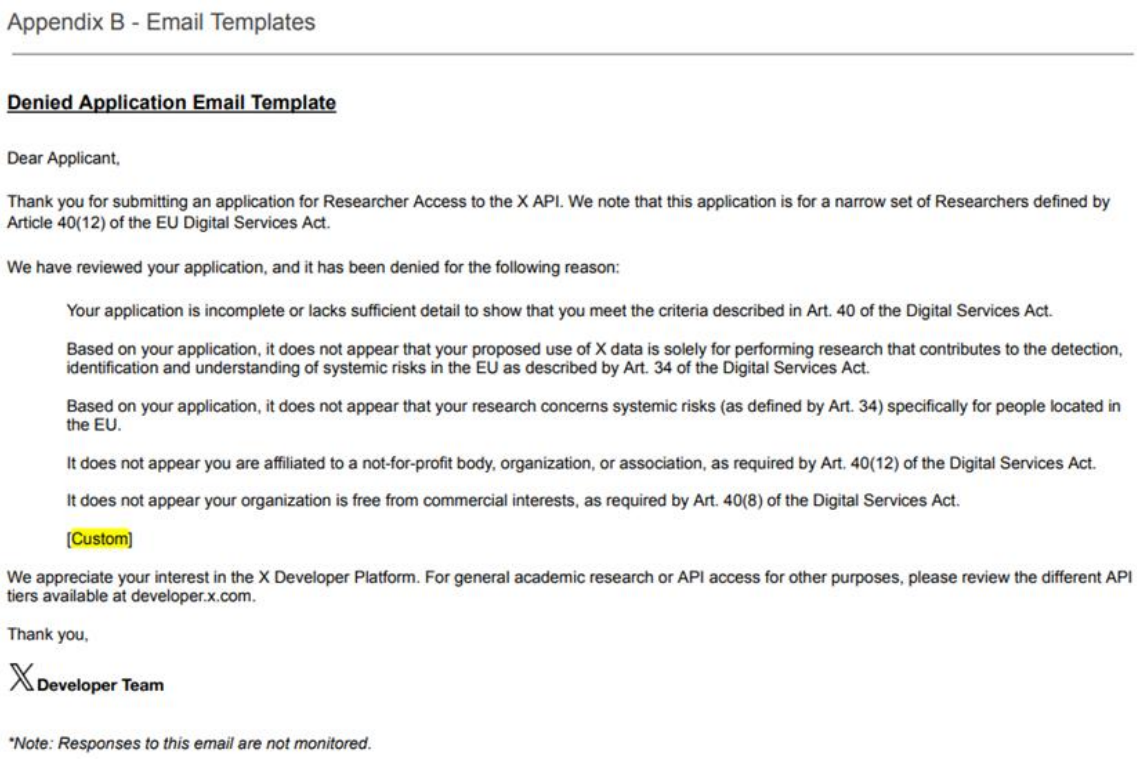
362 As per the table provided in reply to the fourth RFI, request 32)b). This search can be replicated by filtering column R) of the table (DSA.100102, Doc ID 160-2; 160-5).

363 See non-confidential minutes of interview with a researcher, 6 February 2024 (DSA.100102, Doc ID 193-1): '*YY explains that this rejection deters him from applying again as he does not want to waste his time.*'
Non-confidential minutes of interview with researchers, 22 January 2024, (DSA.100102, 40-1): '*When asked by COM if they would try applying again to X's researcher API access, they explain that 'they*

providing meaningful explanations is particularly important where applicants appear to fulfil the criteria set out in Article 40(8), points (b), (c), (d) and (e) of Regulation (EU) 2022/2065 to which Article 40(12) of this Regulation refers.

(340) By failing to properly substantiate the rejections of applications under Article 40(12) of Regulation (EU) 2022/2065, the provider of X has hindered researchers' ability to understand the shortcomings of their applications *alleged* by the provider. In turn, this has undermined researchers' ability to challenge rejections and address potential concerns of the provider of X with the view of obtaining access in the future. Consequently, the failure to meaningfully substantiate rejections undermines the ability of researchers to access data via the X API to perform research on systemic risks in the Union.

Figure 11: Email template used by the provider of X to inform applicants of the denial of their applications.



Appendix B - Email Templates

Denied Application Email Template

Dear Applicant,

Thank you for submitting an application for Researcher Access to the X API. We note that this application is for a narrow set of Researchers defined by Article 40(12) of the EU Digital Services Act.

We have reviewed your application, and it has been denied for the following reason:

Your application is incomplete or lacks sufficient detail to show that you meet the criteria described in Art. 40 of the Digital Services Act.

Based on your application, it does not appear that your proposed use of X data is solely for performing research that contributes to the detection, identification and understanding of systemic risks in the EU as described by Art. 34 of the Digital Services Act.

Based on your application, it does not appear that your research concerns systemic risks (as defined by Art. 34) specifically for people located in the EU.

It does not appear you are affiliated to a not-for-profit body, organization, or association, as required by Art. 40(12) of the Digital Services Act.

It does not appear your organization is free from commercial interests, as required by Art. 40(8) of the Digital Services Act.

[Custom]

We appreciate your interest in the X Developer Platform. For general academic research or API access for other purposes, please review the different API tiers available at developer.x.com.

Thank you,

Developer Team

**Note: Responses to this email are not monitored.*

Source: Annex to reply to the second RFI, request 5 of section VII (DSA.100102, Doc ID 210-8).

6.3.2.3. Insufficient data access quotas & durations for Qualified Researchers

(341) In the third place, the Commission took the preliminary view that the processes devised by the provider of X to obtain data access suppressed the ability of Qualified Researchers to obtain sufficient access duration and access quotas to allow them to perform research that contributes to the detection, identification and understanding of

are not entirely sure the game is fair', also noting that some of their colleagues also had their applications rejected, and that no one they knew had a successful application.'

Non-confidential minutes of interview with a researcher, 6 February 2024 (DSA.100102, 179-1): '*In a recent conference, all researchers he spoke with had been rejected by X. Some researchers he knew explained that they would not apply because they thought it would be a waste of time and resources.'*.

the systemic risks to which X gives rise in the Union, pursuant to Article 40(12) of Regulation (EU) 2022/2065.

(342) The standard access duration and quotas offered by the provider of X are, in fact, substantially inferior to the access that researchers were previously provided under the Twitter Academic API program, to the point of being unusable for some of the research on systemic risks eligible for access pursuant to Article 40(12) of Regulation (EU) 2022/2065.[364] To the knowledge of the Commission, these standard limitations are not conveyed to applicants at any point before or during the application process implemented by the provider of X to grant access to data pursuant to Article 40(12) of Regulation (EU) 2022/2065, but are only communicated at the moment of granting access following a successful application.

6.3.2.3.1. Access durations

(343) First, the provider of X systematically limits successful applicants' API access to a period of 6 months. That is so irrespective of the numerous detailed questions asked by the provider of X during the application process regarding the researchers' data needs and research duration,[365] and irrespective of the fact that the provider of X may have been informed by the researcher that access is needed for a longer period.

(344) As a result, the provider of X systematically provides successful applicants with shorter access periods than requested in their applications,[366] with uncertain extensions.[367] Furthermore, applicants are only informed about applicable data access

---

364 Default access quotas are ten times smaller than under the Twitter Academic program, while default access duration is now 6 months as opposed to indefinite or tailored durations under the Twitter Academic program.; See the following archived webpages: The Internet Archive, Twitter API for Academic Research | Products | Twitter Developer Platform, capture of the website as it appeared on 12 February 2023, accessed on 19 June 2024 (DSA.100102, Doc ID 207-141); The Internet Archive, https://web.archive.org/web/20230212182612/https:/developer.twitter.com/en/use-cases/do-research/academic-research/resources, accessed on 19 June 2024 (DSA.100102, Doc ID 207-145).; GitHub maintained by a Twitter/X developer dating from March 2021: GitHub, https://muhark.github.io/python/scraping/tutorial/2021/03/25/getting-started-with-academic-twitter.html, accessed on 19 June 2024 (DSA.100102, Doc ID 207-125).
See also figures 12 and 13 below.

365 See non-confidential minutes of interview with a researcher, 19 March 2024 (DSA.100102, Doc ID 174-1).

366 The provider of X provides access to successful applicants only for approximately 6 months after their application was accepted, irrespective of the project length for which researchers have applied Out of the ■ successful applicants who obtained access to the X API, ■ decisions to grant access were communicated to applicants on 26 January 2024 and ■ were communicated on 22 April 2024. These two sets of successful applicants were granted access until 30 June 2024 and 31 October 2024 respectively. See the table provided in reply to the fourth RFI, request 32)b) (DSA.100102, Doc ID 160-2; 160-5); See in particular rows which correspond to instances where the provider of X did not provide access for the requested period despite assessing that the applicants fulfilled the conditions to be granted access. See also minutes of interview with a researcher, 19 March 2024 (DSA.100102, Doc ID 174-1): 'YY regrets that he was only provided access until 30 June 2024 without any explanation by X despite requesting access until September 2026 (i.e. the end of his project). YY explained that he intended to resubmit an application to X later in 2024 for his access to continue, but that this would cost him time'.

367 See the table provided in reply to the fourth RFI, request 32)b) (DSA.100102, Doc ID 160-2; 160-5); The Commission notes that the column S of the table provided in reply to the fourth RFI, request 32(b) refers to the 'initial duration' of the access granted to researcher. However, the 'Application Approved' email template (different versions in documents TIUC-1771, 1780, 1798, 1818) ; used by the provider of X to inform researchers of the duration of their access mentions 'we will provide you access to the X

quotas when they are granted access by the provider of X. This means that any follow-up discussion between successful applicants and the provider of X regarding an extension of the quotas happens during that 6-month period during which researchers have been granted access by the provider of X, potentially further reducing the effective access period.[368] This hinders the ability of Qualified Researchers to conduct research on the systemic risks to which X gives rise in the Union by creating uncertainty as to the duration and feasibility of the research. This is particularly burdensome for longitudinal studies,[369] because researchers are required to reapply regularly to the provider of X for data access, without indications from that provider of how that re-application process will work.

6.3.2.3.2. Access quotas

(345) Second, according to researchers interviewed by the Commission, the access provided by default to all Qualified Researchers, namely the 'Pro' API access tier, is insufficient in terms of rate limits and data access quotas to conduct some research projects on systemic risks related to the provision of X in the Union.[370] The 'Pro' tier of the X API allows researchers to access 1 million tweets per month, which is inferior by a magnitude of 10 to the standard access that was formerly provided free of charge to thousands of academics to the Twitter v2 API under the Twitter Academic API program.[371] As reflected in Figures 12 and 13 below, prior to the Spring of 2023, the provider of Twitter was providing more extensive access to data free of charge to thousands of researchers under the Twitter Academic API program (as opposed to only 12 researchers in total who were granted access between 28 August 2023 and 8 May 2024).

(346) Furthermore, to the knowledge of the Commission, the public resources of the provider of X do not mention the existence of this standard quota, fostering an information asymmetry and hindering the ability of Qualified Researchers to tailor

---

Developer API (Pro Tier) until [date]', making no mention that this duration is an initial duration, causing uncertainty for the feasibility of the research. (DSA.100102, Doc ID 168).

368 See for instance the non-confidential minutes of interview with a researcher, 19 March 2024 (DSA.100102, Doc ID 174-1): who has obtained access on 26 January but was still negotiating to obtain larger quotas with X where his initial access was close to its end on 30 June 2024. To the knowledge of the Commission, the provider of X explained this delay only once in its interactions with researchers ("*Additionally, for newly approved researchers with longer term projects, we provide an initial term of access and then reassess on a periodic basis to ensure the access continues to be necessary and proportional to the research. We'll then provide an extension assuming that the researcher still meets the Article 40 criteria.").*; See documents provided in reply to the fourth RFI, request 32)c) (DSA.100102, Doc ID 160-2; 160-5; 168); in particular document TIUC-1085, where a researcher who was granted access asked for follow-up and received information from X.

369 A longitudinal study involves repeated observations of the same variables over long periods of time.

370 See for instance the non-confidential minutes of interviews with researchers on 19 March 2024, 1 March 2024, 28 February 2024 (DSA.100102, Doc ID 174-1; 116-1; 113-1). See the non-confidential testimonies of researchers, 27 February 2024 (DSA.100102, Doc ID 91-1).

371 Figures 12 and 13 provide an overview of the applicable limits on the retrieval of tweets under Twitter/X's past and current data access offerings. Currently, under the Pro API, the limit is at 1M tweets per month, whereas under the free Twitter Academic API program discontinued in Spring 2023, the limit was at 10M per month with no historical restriction, which means that there has been a ten-fold decrease of this limit. For current X API offerings, see https://developer.x.com/en/docs/twitter-api/getting-started/about-twitter-api, accessed on 19 June 2024 (DSA.100102, Doc ID 207-58).For past API offerings, see The Internet Archive, https://web.archive.org/web/20230212083710/https:/developer.twitter.com/en/products/twitter-api/academic-research, accessed on 19 June 2024 (DSA.100102, Doc ID 207-141).

their application to request sufficient data from the beginning of their access period. These practices of the provider, at the very least, add undue delay to the ability of Qualified Researchers to obtain adequate data access quotas to perform research on systemic risks of X in the Union,[372] which is particularly damaging since the standard access quotas provided to successful applicants appear to be insufficient for some types of research on systemic risks in the Union.

(347) Furthermore, evidence from the ■ researchers whose requests were approved by the provider of X indicates that that provider has so far refused to provide higher access quotas even when it was requested by researchers upon the initial approval of their application[373] and where, according to an industry specialist, it would not be complicated nor costly for the provider of X to increase such quotas where needed.[374]

372 Example in non-confidential minutes of interview with a researcher, 19 March 2024 (DSA.100102, Doc ID 174-1): who has obtained access on 26 January but was still negotiating to obtain larger quotas with X where his initial access was close to its end on 30 June 2024. To the knowledge of the Commission, the provider of X explained this delay only once in its interactions with researchers; See documents provided in reply to the fourth RFI, request 32)c) (DSA.100102, Doc ID 160-2; 160-5; 168); in particular document TIUC-1085, where a researcher who was granted access asked for follow-up and received information from X '*Additionally, for newly approved researchers with longer term projects, we provide an initial term of access and then reassess on a periodic basis to ensure the access continues to be necessary and proportional to the research. We'll then provide an extension assuming that the researcher still meets the Article 40 criteria.'.*

373 At the time of the Preliminary Findings, not a single researcher had obtained access to larger quotas than the standard quotas under the 'Pro' access tier, while ■ of the ■ successful applicants explicitly requested such larger quotas following the approval of their application. According to an analysis of the documents provided in reply to the fourth RFI, request 32)b) (DSA.100102, Doc ID 160-2; 160-5; 168-96; 168-110).

374 Non-confidential minutes of interview with an expert, 10 January 2024 (ID 45-1).

Figure 12: Tables featuring the API access tiers available before the spring of 2023. The free of charge Academic Access tier, which allowed researchers to retrieve up to 10 million tweets per month is on the right-hand side.

	Essential	Elevated	Elevated+ (coming soon)	Academic Research
Getting access	Sign up	Apply for additional access within the developer portal	Need more? Sign up for our waitlist	Apply for additional access
Price	Free	Free		Free
Access to Twitter API v2	✓	✓		✓
Access to standard v1.1	✗	✓		✓
Access to premium v1.1	✗	✓		✓
Access to enterprise	✗	✓		✓
Project limits	1 Project	1 Project		1 Project
App limits	1 App per Project	3 Apps per Project		1 App per Project
Tweet caps	Retrieve up to 500k Tweets per month	Retrieve up to 2 million Tweets per month		Retrieve up to 10 million Tweets per month

Source: Public presentation given by Twitter in 2022 (DSA.100102, Doc ID 207-145).

Figure 13: Table featuring the API access tiers available from 9 November 2023 onwards, the 'Pro' Tier priced at USD 5 000 per month allows to retrieve a maximum of 1M posts per month.

	Free	Basic	Pro	Enterprise
Getting access	Get Started	Get Started	Get Started	Get Started
Price	Free	$100/month	$5000/month	
Access to X API v2	✔ (Only Post creation)	✔	✔	
Access to standard v1.1	✔ (Only Media Upload, Help, Rate Limit, and Login with X)	✔ (Only Media Upload, Help, Rate Limit, and Login with X)	✔ (Only Media Upload, Help, Rate Limit, and Login with X)	
Project limits	1 Project	1 Project	1 Project	
App limits	1 App per Project	2 Apps per Project	3 Apps per Project	
Post caps - Post	1,500	3,000	300,000	
Post caps - Pull	✗	10,000	1,000,000	
Filtered stream API	✗	✗	✔	
Access to full-archive search	✗	✗	✔	
Access to Ads API	✔	✔	✔	

Source: X Developer Platform, https://developer.x.com/en/docs/twitter-api/getting-started/about-twitter-api, accessed on 19 June 2024 (DSA.100102, Doc ID 207-58).

6.3.2.4. Unnecessary obstacles in the processes set by the provider of X to grant access to data pursuant to Article 40(12) of Regulation (EU) 2022/2065

(348) In the fourth place, several decisions of the provider related to its data access processes add unnecessary obstacles for researchers to apply and obtain access to publicly accessible data via the X API pursuant to Article 40(12) of Regulation (EU) 2022/2065.

6.3.2.4.1. Degradation of the information available to researchers about the possibility to access data

(349) First, the information made available by the provider of X for prospective applicants is considerably degraded compared to the information made available under the Twitter Academic API program until the Spring of 2023. This is despite the fact that the current X API and the Twitter v2 API are almost identical from a technical perspective, meaning that updating these resources would require minimal work. The provider of X/Twitter previously maintained public pages containing comprehensive information for researchers, including a simple description of eligibility conditions as well as the data available to academics. The provider of X/Twitter also provided access to extensive tutorials for research use cases, as well as links to third party resources such as relevant Python Libraries for academic research based on X/Twitter data.[375]

(350) After the Spring of 2023, however, the provider of X took down most of the public pages containing these resources, even when they would still be relevant to support Qualified Researchers under the current program, and the remaining pages appear to be scarcely maintained.

(351) The possibility to apply for access to data pursuant to Article 40(12) of Regulation (EU) 2022/2065 is under-referenced in X's online pages for prospective researchers. At present, the provider of X maintains two information pages for researchers interested in conducting research based on X data (the 'Do Research' page and the 'Academic Research' page[376]) which include considerably less information than in early 2023,[377] when the provider of X terminated its Twitter Academic API program. In particular, the 'Academic Research' page only contains a 'dead' link to an academic research project and a link to X's now mostly inactive academic forums. The 'Do Research' information page features very little information for researchers apart from a link to apply via the 40(12) application form and a reference to three old research projects conducted under Twitter's now-discontinued Twitter Academic API program.

---

375 Examples : Archived Twitter page, https://web.archive.org/web/20230212182612/https:/developer.twitter.com/en/use-cases/do-research/academic-research/resources, accessed on 19 June (DSA.100102, Doc ID 207-145); Archived Twitter page, https://web.archive.org/web/20230212083710/https:/developer.twitter.com/en/products/twitter-api/academic-research, accessed on 19 June (DSA.100102, Doc ID 207-141).

376 'Do Research' https://developer.x.com/en/use-cases/do-research and 'Academic Research' https://developer.x.com/en/use-cases/do-research/academic-research, accessed on 19 June 2024 (DSA.100102, Doc ID 207-62).

377 Examples : Archived Twitter page, https://web.archive.org/web/20230212182612/https:/developer.twitter.com/en/use-cases/do-research/academic-research/resources, accessed on 19 June (DSA.100102, Doc ID 207-145); Archived Twitter page, https://web.archive.org/web/20230212083710/https:/developer.twitter.com/en/products/twitter-api/academic-research, accessed on 19 June (DSA.100102, Doc ID 207-141).

(352) The Commission also notes that, before Regulation (EU) 2022/2065 started applying to X, staff of the provider of Twitter regularly attended academic conferences to present the API in great detail to academic researchers, in addition to making free courses available.[378] To the knowledge of the Commission, the provider of X has not engaged in similar outreach exercises since Regulation (EU) 2022/2065 started applying to X.

6.3.2.4.2. Unnecessary friction resulting from the application form

(353) Second, the application form put in place by the provider of X to fulfil its obligations under Article 40(12) of Regulation (EU) 2022/2065 is designed and implemented in a manner creating unnecessary friction for applicants. The provider of X is at least partially aware of this friction, yet has not taken any steps to effectively reduce it.

(354) Prior to the Spring of 2023, the provider of X allowed researchers to apply directly from their individual free Twitter developer account for access to the Twitter Academic API program.[379] At present, researchers may only submit their applications to the X API via a Google form.[380] This form and the accompanying design choices by the provider of X have several shortcomings that de facto restrict researchers' access to public data under Article 40(12) of Regulation (EU) 2022/2065.

(355) Several researchers interviewed by the Commission services complained that they had no possibility to retrieve their application once it had been submitted via the dedicated Google form set up by the provider of X. One researcher was able to obtain a copy of an application from the provider of X only fifteen days after requesting it, while several other researchers requested copies of their application from the provider of X without success.[381] The Commission notes that Google Forms provide an option to automatically 'send responders a copy of their response',[382] but the provider of X did

---

378 Examples: YouTube, https://www.youtube.com/watch?v=Uv4pGP1jZxE, accessed on 19 June 2024 (DSA.100102, Doc ID 207-149); GitHub repository, https://muhark.github.io/python/scraping/tutorial/2021/03/25/getting-started-with-academic-twitter.html, accessed on 19 June 2024 (DSA.100102, Doc ID 207-125).

379 Overviews of the old process to apply for Twitter API Academic Access: DEV Community, https://dev.to/suhemparack/a-guide-to-teaching-with-the-twitter-api-v2-3n08, accessed on 19 June 2024 (DSA.100102, Doc ID 207-121); YouTube, https://www.youtube.com/watch?v=Uv4pGP1jZxE, accessed on 19 June 2024 (DSA.100102, Doc ID 207-149); Archived Twitter Page, https://web.archive.org/web/20230212182612/https:/developer.twitter.com/en/use-cases/do-research/academic-research/resources, accessed on 19 June 2024 (DSA.100102, Doc ID 207-145).

380 See https://docs.google.com/forms/d/e/1FAIpQLSdo0O-D6Kxa3cV4g1JLz2T_0Sk3hdEnTdv8dJmibagCnzJ7kg/viewform, accessed on 19 June 2024 (DSA.100102, Doc ID 207-70).

381 In one of the cases brought to the attention of the Commission, one researcher asked for a copy of his application on 6 December 2023 and received it 13 days later, on 19 December 2023; X was therefore aware of this issue as early as December 2023 and did not change the setting in order for applicants to automatically receive a copy of their application. See non-confidential minutes of interview of researcher, 26 February 2024 (DSA.100102, Doc ID 86-1) as well as documents TIUC-274, 1438 and 1649 provided in reply to request 32)c) and 34 of the fourth RFI (DSA.100102, Doc ID 160-2; 160-5; 168).
In other instances, researchers requested that a copy of their application be provided, and never obtained a reply from X. See documents TIUC-002, 1231, 1260 provided in reply to the fourth RFI, request 32)c) (DSA.100102, Doc ID 160-2; 160-5; 168).

382 For an overview of the process to enable respondents to a Google Form to obtain a copy of their submission: https://support.google.com/docs/thread/202090575/google-forms-copy-of-responses-not-received?hl=en#:~:text=If%20you%20selected%20When%20requested,not%20get%20a%20response%

not enable that option. Researchers have further explained, in this respect, that the provider of X's practice of not sharing a copy of the research applications by default has detrimental consequences on the ability of researchers to build upon their first application if and when the provider of X asks for additional information at a later stage (which is part of the provider of X's standard review process for applications that pass the first screening).[383] One researcher also regretted that the current form used by the provider of X did not allow it to upload supporting files, such as a CV. In addition, the provider of X does not[384] send an acknowledgment of receipt by email when researchers submit their applications and does not provide researchers with a copy of their application. Instead, that provider only sends researchers a first email around 7 days after receiving their application, which is either a rejection of the request or a notification that the request is under review. [385]

(356) In the Preliminary Findings, the Commission took the preliminary view that the absence of such basic insights into the application process suggests, at the very least, a lack of mechanisms for the provider of X to take researcher feedback into account and iteratively improve its process to give access to data without undue delay in accordance with its obligations under Article 40(12) of Regulation (EU) 2022/2065.

6.3.2.5. Discouraging effect on researchers

(357) Finally, in the Preliminary Findings, the Commission took the preliminary view that the cumulative effects of the shortcomings mentioned in sections 6.3.2.1 to 6.3.2.4 above have *not only* overly restricted access for the [redacted] researchers who applied through the mechanism implemented by the provider of X on 9 November 2023, but have also functioned as a deterrent and discouraging factor to researchers with regards to applying to obtain access to data pursuant to Article 40(12) of Regulation (EU) 2022/2065.

(358) Researchers that were interviewed by the Commission in the context of its investigation explained that the new data access program implemented by the provider of X was widely perceived by the research community as opaque and having a very high number of rejections, meaning that they had very low chances of being approved.[386] Several researchers explained to the Commission that they did not intend to apply again under the current program, while others explained that they had not even tried applying after hearing from colleagues about the opacity of the process and the very high rate of rejections.[387] The actions of the provider of X thus *de facto*

20receipt and https://www.youtube.com/watch?v=eSAfPZRJ3CU, accessed on 19 June 2024 (DSA.100102, Doc ID 207-77; 207-81).

383 See non-confidential minutes of interviews with researchers, 19 March 2024, 22 January 2024, 8 February 2024 (DSA.100102, Doc ID 174-1; 40-1; 133-1).

384 See for instance the non-confidential minutes of the interviews of researchers, 8 February 2024, 22 January 2024 (DSA.100102, Doc ID 133-1; 36-1).

385 See for instance the dates in the table provided in reply to the fourth RFI, request 32)b) (DSA.100102, Doc ID 160-2; 160-5).

386 See for instance non-confidential minutes of interview with a researcher, 6 February 2024 (DSA.100102, 179-1): '*In a recent conference, all researchers he spoke with had been rejected by X. Some researchers he knew explained that they would not apply because they thought it would be a waste of time and resources.'*.

387 See non-confidential minutes of interview with a researcher, 6 February 2024 (DSA.100102, Doc ID 193-1): '*YY explains that this rejection deters him from applying again as he does not want to waste his time*.'
Non-confidential minutes of interview with researchers, 22 January 2024, (DSA.100102, 40-1): '*When asked by COM if they would try applying again to X's researcher API access, they explain that 'they*

deterred researchers from even attempting to obtain access to X's API to conduct research on systemic risks in the Union pursuant to Article 40(12) of Regulation (EU) 2022/2065.

6.3.3. Prohibiting independent access to data during the entire relevant period

(359) In the Preliminary Findings, the Commission took the preliminary view that the provider of X, by largely prohibiting Qualified Researchers from *independently* accessing X's publicly accessible data, including by automated means such as data crawling and data scraping, has contributed to prevent Qualified Researchers from accessing and using data publicly accessible in X's online interface to perform research contributing to the detection, identification and understanding of the systemic risks within the meaning of Article 34(1) of Regulation (EU) 2022/2065, in contravention of Article 40(12) of that Regulation.

(360) To comply with its obligations under Article 40(12) of Regulation (EU) 2022/2065, the provider of X should not only give Qualified Researchers access to data publicly accessible in X's online interface by means of the X API, but should also refrain from prohibiting Qualified Researchers from independently accessing X's publicly available data, including by means of data scraping. This is clarified in recital 98 of Regulation (EU) 2022/2065, which states that '*where data is publicly accessible, […] providers [of very large online platforms or of very large online search engines] should not prevent researchers meeting an appropriate subset of criteria from using this data for research purposes that contribute to the detection, identification and understanding of systemic risks*' (emphasis added).

(361) The Commission noted that independent access techniques, including automated techniques such as scraping and crawling, are crucial to enable Qualified Researchers to perform specific types of research on systemic risks in the Union within the meaning of Article 34(1) of Regulation (EU) 2022/2065. For instance, independent access is necessary to perform research into the design and functioning of X's recommender systems.[388] In addition, independent access to data is necessary to audit the quality of the data provided through the X API.

(362) In response to the Commission's request to provide additional information regarding its relevant terms of service and policies in relation to independent access to publicly accessible data under Article 40(12) of Regulation (EU) 2022/2065, the provider of X stated that '*Section 4 of X's Terms of Service imposes restrictions on misusing the service*'.[389] Section 4 of X's terms of service stipulate the following: '*You may not do any of the following while accessing or using the Services: […] (iii) access or search or attempt to access or search the Services by any means (automated or otherwise) other than through our currently available, published interfaces that are provided by us (and only pursuant to the applicable terms and conditions), unless you have been*

---

are not entirely sure the game is fair', also noting that some of their colleagues also had their applications rejected, and that no one they knew had a successful application.'
See also the non-confidential minutes of interviews with researchers, 8 February 2024, 6 February 2024, 1 March 2024 (DSA.100102, 133-1; 116-1).

388 For instance, independent access to publicly accessible data enables Qualified Researchers to test how certain actions influence user timelines mediated by recommender systems, which can be an important factor influencing systemic risks. Article 34(2)(a) of Regulation (EU) 2022/2065 unequivocally references '*the design of [their] recommender systems*' as an influence factor on systemic risks. The X API currently does not allow this type of research.

389 See the reply to the second RFI, request 3 of section VII (DSA.100102, Doc ID 210-9).

specifically allowed to do so in a separate agreement with us (NOTE: crawling or scraping the Services in any form, for any purpose without our prior written consent is expressly prohibited);'.[390] In the same reply, the provider of X also explained to the Commission that *'Researchers who access data by either subscribing to a Developer API access tier or by applying via the Art.40(12) Application form are not bound by this restriction in X's Terms of Service, but instead are bound by the terms in the X Developer Agreement*'. The Commission notes that the 40(12) application form was implemented by the provider of X on 9 November 2023 and that the X Developer Agreement mentioned was updated on 15 November 2023 to include language related to Article 40 of Regulation (EU) 2022/2065.[391]

(363) Therefore, from 28 August to 9 November 2023, the only contractual terms relevant to Qualified Researchers' independent access to X's publicly accessible data was section 4 of X's Terms of Service, which includes a general prohibition of independent access to data[392] without prior written approval of the provider of X. X's terms of service do not include any exception to this general prohibition, which is therefore indiscriminately applicable to all recipients of X, including Qualified Researchers. During this period, the provider of X had no effective process in place for Qualified Researchers to obtain the prior written approval referred to in X's terms of service,[393] effectively preventing such researchers from independently accessing X's publicly accessible data without infringing those terms of service, in contravention of Article 40(12) of Regulation (EU) 2022/2065.

(364) On 15 November 2023, the provider of X modified the language of its Developer Agreement by adding a reference to 'Article 40 of [Regulation (EU) 2022/2065]'. [394] However, the Commission considered that this addition does not provide Qualified Researchers with sufficient clarity regarding their ability to independently access data.

---

390 See X's terms of service: https://x.com/en/tos, accessed on 19 June 2024 (DSA.100102, Doc ID 207-46).

391 Open Terms Archive, https://github.com/OpenTermsArchive/pga-versions/commit/c63b906e573f9f103649430a31a8080d2f7544e3#diff-2d91701cdd46e3d6db2a5cac7bd6531c806a0723a4f01ca8a1a6075eb98b533aR90, accessed on 19 June 2024 (DSA.100102, Doc ID 207-153).
The added language in the X Developer agreement stipulates the following: '***Digital Services Act***. *Notwithstanding anything to the contrary in this Agreement, to the extent you are provided access to the Licensed Material pursuant to the procedures described in Article 40 of the Digital Services Act (The Digital Services Act) ('DSA'), your access and use of the Licensed Material is limited solely to performing research that contributes to the detection, identification and understanding of systemic risks in the European Union and only to the extent necessary for X to comply with its obligations under the DSA. Any such use of the Licensed Material is non-commercial as described in Section III(B) of this Agreement. You may not disclose, reproduce, license, or otherwise distribute the Licensed Material (including any derivatives thereof) that you retrieve through the X API to any person or entity outside the persons within your organization necessary to perform the research, unless (i) the information is disclosed to the Digital Services Coordinator or other party specifically permitted by the DSA pursuant to the 'vetted researcher' status and procedures described in Article 40, or (ii) disclosure is required by law.*'.

392 The Commission notes that this general prohibition of accessing publicly accessible data in X's online interface includes prohibiting Qualified Researchers from manually (i.e. by means of copy pasting) or automatically extracting necessary passages of text or media data from the X online interface.

393 In line with the elements put forward in section 6.3.1. above.

394 Open Terms Archive, https://github.com/OpenTermsArchive/pga-versions/commit/c63b906e573f9f103649430a31a8080d2f7544e3#diff-2d91701cdd46e3d6db2a5cac7bd6531c806a0723a4f01ca8a1a6075eb98b533aR90, accessed on 19 June 2024 (DSA.100102, Doc ID 207-153).

Since the provider of X has not amended the wording prohibiting independent access in section 4 of its terms of service, it maintains an ambiguity as regards the prohibition of independent access for Qualified Researchers, who might not fall under the Developer Agreement. The provider of X has not addressed this ambiguity despite the Initiation of Proceedings Decision explicitly referring to the prohibition of scraping by Qualified Researchers in X's terms of service constituting a suspected infringement of Article 40(12) of Regulation (EU) 2022/2065.

(365) This persistent ambiguity in X's terms and conditions[395] in relation to independent access has led to negative consequences for Qualified Researchers. In particular, research projects relying on independent access are unlikely to receive a positive opinion from institutional review boards or from academic journals, given the contractual violation such access would present. [396]

(366) In light of the foregoing, the Commission took the preliminary view that the provider of X prevents Qualified Researchers from independently accessing X's publicly accessible data and thereby undermined their ability to perform research that contributes to the detection, identification and understanding of systemic risks in the Union within the meaning of Article 34(1) of Regulation (EU) 2022/2065, in contravention of Article 40(12) of that Regulation.

6.4. The arguments of the provider of X

(367) In the Reply to the Preliminary Findings, the provider of X challenges the Commission's preliminary findings in relation to its alleged non-compliance with Article 40(12) of Regulation (EU) 2022/2065 in the Union on several grounds which the Commission has summarised into seven categories.

(368) In the first place, the provider of X claims that, in the absence of a Commission delegated act and guidance, it has struck a reasonable balance between its legitimate interests and other applicable rules when assessing data requests by researchers.

(369) First, the provider of X notes that access to publicly accessible data pursuant to Article 40(12) of Regulation (EU) 2022/2065 is conditional on applicants meeting five cumulative requirements, namely the four conditions set out in Article 40(8), points (b), (c), (d), and (e) of that Regulation and the requirement to use the data solely for performing research that contributes to the detection, identification and understanding of systemic risks in the Union pursuant to Article 34(1) of Regulation (EU) 2022/2065.[397] According to the provider of X, the '*PF's narrative seems to suggest*

---

395 For the purposes of this document, 'terms and conditions' mean all clauses, irrespective of their name or form, which govern the contractual relationship between the provider of intermediary services and the recipients of the service, in line with Article 3(u) of Regulation (EU) 2022/2065.

396 See non-confidential minutes of interviews of researchers, 18 March 2024 (DSA.100102, Doc ID 122-1): '*YY is reluctant to scrape as it currently means that his research would be going against X's terms of service, causing ethical review boards to be reluctant to approve his research. When asked to clarify, YY explains that X's terms of service seem to prohibit scraping, and that ethical review boards are unlikely to approve research projects that violate platforms' terms*'.
Non-confidential minutes of interview with an expert, 10 January 2024 (DSA.100102, Doc ID 45-1): '*He noted that researchers are also dissuaded by X to collect data via other means, such as scraping'*.
See also application 23-128, contained in the table and documents provided in reply to the fourth RFI, request 32)b) (DSA.100102, Doc ID 160-2; 160-5; 168): In cell J132, the applicant justified its request by explaining that '*This data is currently not available to us as the academic API is not functional anymore and the X platform prohibits scraping of its content.'.*

397 Reply to the Preliminary Findings, paragraph 245.

that access to data under Article 40(12) DSA shall be granted to applicants almost automatically, overlooking the applicability of those requirements.'[398]

(370) Second, the provider of X argues that '*the Commission and several Member States have failed to adopt the necessary measures for researchers to be able to seek access to data pursuant to these provisions*', meaning both Article 40(4) and Article 40(12) of Regulation (EU) 2022/2065. According to the provider of X, the '*Commission has failed to adopt the delegated acts required for this regime to be enforceable, and several Member States have failed to appoint Digital Services Coordinators*'.[399] The provider of X further claims that '*the Commission has always been well aware of the need to specify and clarify the content of Article 40 DSA and, therefore, expressly envisaged, already in the DSA proposal, that the Commission 'shall' adopt delegated acts laying down the specific conditions governing the sharing of data with researchers. The EU legislator retained this wording in the final text*', namely in Article 40(13) of Regulation (EU) 2022/2065.[400] According to the provider of X, Article 40(13) '*imposes an unequivocal mandate on the Commission*' to adopt delegated acts, which the Commission has not yet done,[401] '*choosing instead to prioritize bringing and publicizing an unwarranted case against X'*.[402] The provider of X argues that the delegated act is also '*applicable to access requirements under Article 40(12) DSA, since they are virtually the same as those applicable to vetted researchers under Article 40(4). In particular, the requirements set out in points (b), (c), (d) and (e) of Article 40(8) DSA are indistinctly applicable to both vetted and Qualified Researchers*'.[403]

(371) The provider of X also interprets Article 93 of Regulation (EU) 2022/2065[404] as mandating the adoption of delegated acts pursuant to Article 40(13) of that regulation by 16 November 2022 and accuses the Commission of being '*in breach of its legislative duties […] at the very least since that date*'.[405] For the provider of X, the '*necessary consequence of the Commission's failure to comply with its obligation under Article 40(13) DSA is that the DSA data access regime remains not only unclear, but also incomplete and not fully functional or unenforceable'*.[406]

(372) Third, the provider of X argues that the eligibility requirements under Article 40(12) of Regulation (EU) 2022/2065 are unclear, meaning that, for the data access regime to operate as intended, the Commission must provide clarity and guidance, notably as regards its interaction with Regulation (EU) 2016/679. Without such guidance, the provider of X considers that its approach has been '*prudent and lawful*'.[407] The

---

398 Reply to the Preliminary Findings, paragraph 246.

399 Reply to the Preliminary Findings, paragraph 240.

400 Reply to the Preliminary Findings, paragraph 248 and 249.

401 The Commission published the draft delegated regulation on 28 October 2024 for public consultation; https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act_en. The provider of X did contribute to this public consultation on 10 December 2024 : https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act/F3499020_en.

402 Reply to the Preliminary Findings, paragraph 249 and 250.

403 Reply to the Preliminary Findings, paragraph 251.

404 Article 93 of Regulation (EU) 2022/2065 mentions that '*Article 40(13) […] shall apply from 16 November 2022'*.

405 Reply to the Preliminary Findings, paragraph 252.

406 Reply to the Preliminary Findings, paragraph 253.

407 Reply to the Preliminary Findings, paragraph 256.

provider of X claims that any other course of action could '*seriously impact the fundamental rights of users who have placed their trust in X*', and states that it has interpreted Article 40(12) of Regulation (EU) 2022/2065 in the light of recital 97 of that regulation, which mentions that '*[a]ll requests for access to data under that framework should be proportionate and appropriately protect the rights and legitimate interests, including the protection of personal data, trade secrets and other confidential information, of the very large online platform or of the very large online search engine and any other parties concerned, including the recipients of the service*.'[408] The provider of X argues that a lax application of the requirement could lead to granting access to '*researchers who, under the pretext of seeking to analyze systemic risks in the Union stemming from the design or functioning of a given VLOP, may in reality wish to gather information to use (or to share) with businesses or hostile governments seeking to capitalize on information about X users or to manipulate public opinion during an election campaign*', which would counter the DSA's purpose.[409] The provider of X claims that this alleged lack of legal clarity places very large online platforms in a difficult position regarding compliance with the GDPR and is incompatible with the general principle of legal clarity.[410]

(373) The provider of X then argues that '*even publicly accessible data can reveal information and insights that may be of a sensitive nature*', citing posts deleted by users or because of a court order, as well as insights derived from the analysis of large volumes of data.[411] The provider of X also notes that publicly accessible data can be used by third parties to gain significant commercial or strategic advantages, referring to Directive (EU) 2016/943 ('the Trade Secrets Directive'). The provider of X then points to World Trade Organisation (WTO) rules, which establish that the '*protection against unfair competition*' is a lawful reason for preventing information from being disclosed.[412] The provider of X claims that these considerations are the reason for which Article 40(12) of Regulation (EU) 2022/2065 requires that access be granted only to researchers *'capable of fulfilling the specific data security and confidentiality requirements corresponding to each request*'.[413]

(374) In the second place, the provider of X claims to have put two solutions in place from 28 August 2023 to 9 November 2023 to provide Qualified Researchers with access to its publicly available data pursuant to Article 40(12) of Regulation EU 2022/2065. The first solution consisted in offering the possibility to subscribe to X's commercial 'developer API' service.[414] The second solution consisted in the possibility for researchers to request access via the email address EU-Questions@x.com, which is published on the EU-specific section of X's Help Center and presented as the general

408 Reply to the Preliminary Findings, paragraph 257.
409 Reply to the Preliminary Findings, paragraph 258.
410 Reply to the Preliminary Findings, paragraph 261.
411 Reply to the Preliminary Findings, paragraph 263 and 264.
412 In particular, Article 39 TRIPS, which *'allows natural and legal persons "the possibility of preventing information lawfully within their control from being disclosed [...] without their consent in a manner contrary to honest commercial practices" so as to ensure "effective protection against unfair competition".*'; World Trade Organization Agreement on Trade-Related Aspects of Intellectual Property Rights of 15 April 1994 ('TRIPS'), TRT/WTO01/001.
413 Reply to the Preliminary Findings, paragraph 265.
414 The 'X Developer API' ('X API') can be used to programmatically retrieve and analyse X data, it is accessible to any individual or entity paying the costs set by the provider of X and accepting the relevant terms of service.; see https://developer.x.com/en, accessed on 21 June 2024 (DSA.100102, Doc ID 207-165).

point of contact for all 'additional queries' that cannot be resolved only by consulting the Help Center.[415]

(375) First, the provider of X insists that the aforementioned email address was displayed on X's interface and notes that the EU-Specific section of its Help Center '*displays, prominently, a link to the Commission's dedicated official DSA website explaining any and all rights arising out of the DSA, including in relation to researcher access to data*'.[416] This email address was presented by the provider of X as the general point of contact for all '*additional queries*' that cannot be resolved only by consulting the Help Center.[417] The provider of X also claims that, over this period, it provided '*guidance for researchers, including researchers' communities and tailored get starting guides*'.

(376) Second, the provider of X argues that the absence of a documented, systematic process to review applications does not imply that a request would not have been handled correctly.[418] The provider of X claims that if researchers had asked for access through this channel and had been successful in obtaining such access, the provider of X would have given such access for free.[419]

(377) Third, according to the provider of X, payment for access to data for research purposes is so widespread and accepted that usually specific funding is granted for such purposes.[420] The provider of X references one example of such a grant.[421]

(378) In the third place, the provider of X claims that its interpretation of the applicable requirements for data access was the correct one.

(379) First, the provider of X disagrees with the Commission's interpretation set out in recital 321 above, emphasising that '*the use of the term solely emphasises exclusivity in purpose is legitimate (i.e. the use of the data shall directly and exclusively contribute to the detection, identification and understanding of systemic risks within the Union*)' (emphasis added).[422] The provider of X argues that the Commission's interpretation departs from the wording of Article 40(12) of Regulation (EU) 2022/2065, arguing that the legislative intent behind using the word *solely* is to restrict the use of data access under that provision only to specific purposes that contribute to the understanding of systemic risks. According to the provider of X, '*the use of the term 'contributes' supports this interpretation and imposes an additional requirement that the use of the data must be useful or effective in achieving the intended purpose. For this reason, any data usage under this provision needs to demonstrably advance the purpose of the research, ensuring that activities are not only compliant in form but also substantive in contributing to the intended outcomes. This requirement adds a layer of accountability, compelling researchers to demonstrate the practical relevance and impact of their data usage, aligning it directly with the regulatory goals*'

---

415 This information was communicated to the Commission on 17 November in the reply to the first RFI, section 5 in parallel to the launch of the new application process discussed in paragraph 294 (DSA.100102, Doc ID 2-11); The provider of X reiterated this information in the Reply to the Preliminary Findings, paragraph 266.

416 Reply to the Preliminary Findings, paragraph 266.

417 Reply to the Preliminary Findings, paragraph 266.

418 Reply to the Preliminary Findings, paragraph 323.

419 Reply to the Preliminary Findings, paragraph 321.

420 Reply to the Preliminary Findings, paragraph 321.

421 Reply to the Preliminary Findings, footnote 378 - University of [redacted], '[redacted] Grant', available at: [redacted]

422 Reply to the Preliminary Findings, paragraph 285.

(emphasis added).[423] The provider of X emphasises that its '*cautious interpretation*' of the term solely is even more important with regards to preventing misuse of personal and confidential information contained in the data.

(380) The provider of X further claims that the Commission did not *'conduct any assessment of whether non-accepted applications should in fact have been accepted*'.[424] The provider of X also argues that the in-depth discussion of rejections provided in the Commission's technical analysis is '*incapable of supporting any preliminary finding of infringement*'.[425] Firstly, the provider of X argues that the sample of applications analysed in-depth was too small to be relevant, and had no representative value since the Commission had chosen these applications '*where it appeared that processes and reasoning from the provider to deny applications were particularly flawed*' (quoting the technical analysis).[426] Secondly, the provider of X challenges the observation that these specific applications were wrongfully denied.

(381) The provider of X criticises the validity of the survey report published by the European New School and the Weizenbaum Institute, which criticised that provider's interpretation of the scope of eligible research projects as being too narrow. [427] The provider of X argues that this report does not support the view that it would have systematically failed to comply with Article 40(12) for three reasons.[428] Firstly, the report relies on a tracker to which researchers could voluntarily submit data about their applications to different providers pursuant to Article 40(12) and was not intended to serve as evidence for systematic conclusions about possible breaches of the Regulation. The provider of X also claims that the study does not consider whether the applications themselves had deficiencies. Secondly, the report does not disclose its sample size and the distribution of applications across providers and is likely to suffer from selection bias due to its reliance on crowdsourcing. Thirdly, the report was intended to analyse trends on the early implementation of the Regulation, and to highlight points for improvements, one of which was to encourage the Commission to issue delegated acts pursuant to 40(13) and develop guidelines for Article 40(12).

(382) Second, the provider of X claims that it has not systemically rejected non-EU researchers based solely on their location, giving the example of application [redacted] where an EU researcher was allegedly granted access in relation to an application including other non-EU researchers.[429] The provider of X claims that the [redacted] applications referenced by the Commission as rejected on the basis that the applicants were established outside the Union were actually rejected because of provider's standard review processes to ensure that the proposed research has an '*EU Nexus*' (in the provider's words).[430]

---

423 Reply to the Preliminary Findings, paragraph 286.

424 Reply to the Preliminary Findings, paragraph 277.

425 Reply to the Preliminary Findings, paragraph 279.

426 Reply to the Preliminary Findings, paragraph 280; 'DSA.100102' Technical Analysis (DSA.100102, Doc ID 257-1), paragraph 2.

427 Weizenbaum Institute, https://www.weizenbaum-library.de/items/39a490c3-f3c1-42a7-a530-318b17e9de49, April 2024, in particular 'Sample Case 1' on Page 5, and page 6 of the report (DSA.100102, Doc ID 128-1): '*it is an unnecessarily narrow interpretation of the provision to restrict data access by claiming a request is not specific enough without further explanations*'.

428 Reply to the Preliminary Findings, paragraph 278.

429 Reply to the Preliminary Findings, paragraph 288.

430 Reply to the Preliminary Findings, paragraph 287.

(383) In parallel, the provider of X argues that the legal text supports its interpretation of Article 40(12) of Regulation (EU) 2022/2065 that it is justified to reject researchers solely because they are established or located outside the Union.[431] For that provider, the fact that Regulation (EU) 2022/2065 applies to intermediary services offered to recipients established in the Union implies that its data access provisions are limited to researchers established in the Union. The provider of X further argues that EU-based research organisations should in any case be favoured, in order to promote research on systemic risks in the Union and to ensure that research complies with EU standards for data protection.[432] In particular, that provider insists on its intention to minimise risks arising out of international data transfers.[433] The provider of X also claims that it took its decision to restrict access to EU-based organisations in the absence of guidance in the form of a delegated regulation from the Commission, and argues that its interpretation is shared by other stakeholders, quoting mainly other providers and texts that were published prior to and during the call for evidence on the delegated acts pursuant to Article 40(13) of Regulation (EU) 2022/2065 launched in May 2023, which pre-dates the entry into force of that Regulation.[434]

(384) Third, the provider of X argues that the Commission does not put forward evidence that it may have rejected in practice researchers who are not affiliated with a not-for-profit body, organisation or association.[435] The provider of X also claims that the specific reference to '*researchers affiliated to non-for-profit bodies, organizations, and associations'* is indicative of a deliberate choice by the EU legislature to exclude researchers who are not affiliated to any not-for-profit body, organisation or association.

(385) Fourth, the provider of X argues that, in absence of guidance from the Commission regarding the requirements set out in Article 40(8), points (b) and (c) of Regulation (EU) 2022/2065,[436] should some fields of its application not be strictly necessary in the Commission's view to carry out the assessment of the requirements of Article 40(12) of that Regulation, it would not constitute an element of non-compliance. [437] In relation to the mandatory '*Commercial Interest*' field of its application form, the provider of X argues that *'X's application form includes a list of elements that, although not mandatory, may be useful for researchers in order to demonstrate that 'they are independent from commercial interest', as required by Article 40(8), point (b)'*.[438] In relation to the requirement for applicants to disclose '*the funding of [their] research, including amounts raised and direct and indirect sources of funding*', the provider of X argues that it is following the trends in other EU regulations, which

431 Reply to the Preliminary Findings, paragraph 289.
432 Reply to the Preliminary Findings, paragraph 289.
433 Reply to the Preliminary Findings, paragraph 290.
434 EC Call for evidence on the Delegated Regulation on data access provided for in the Digital Services Act (Spring 2023), https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act_en.
Summary of the contributions to the Call for evidence on the Delegated Regulation on data access provided for in the Digital Services Act https://ec.europa.eu/newsroom/dae/redirection/document/100331, accessed on 17 June 2025 (DSA.100102, Doc ID 366); The provider of X did not submit a direct contribution to this call for evidence.
435 Reply to the Preliminary Findings, paragraph 297.
436 Reply to the Preliminary Findings, paragraph 293.
437 Reply to the Preliminary Findings, paragraph 294.
438 Reply to the Preliminary Findings, paragraph 295.

impose strict funding transparency requirements, such as regulations to counter money laundering or on foreign subsidies. The provider of X claims to be open to adapting its process based on Commission guidance, but that no infringement of Regulation (EU) 2022/2065 can be established on the basis of this specific conduct. [439]

(386) In the fourth place, the provider of X claims that its process to assess applications and grant Qualified Researchers access to data did not suffer from significant shortcomings after 9 November 2023.

(387) First, the provider of X claims to always indicate to rejected applicants which specific requirement(s) of Article 40(12) of Regulation (EU) 2022/2065 their application does not meet. In particular, that provider explains that '*to further streamline the process, ensure a timely and effective assessment and reduce any risk of 'undue delay', X set up a staggered approach procedure entailing a first review of the applications as part of which only those which 'clearly' did not meet the Article 40(12) DSA requirements were rejected. All other applications were considered to be potentially eligible for access; consequently, X's Developer Team engaged in a dialogue with the researchers at issue to verify that they fulfilled the relevant requirements*'.[440] The provider of X also claims that its processes *'usually involves a back-and forth of emails with applicants to explain what the application is lacking and to assist them throughout the application process'.* That provider further argues that it has '*provided all applicants with an explanation for why the application was not accepted, and the Commission has failed to explain why this explanation is insufficient*'. That provider also argues that '*[n]owhere in the DSA is it stated that VLOPs have to provide researchers with more detailed grounds for the rejection of their application*'. [441]

(388) Second, the provider of X claims that its processes, including prior to January 2024, exceed what is legally required under Article 40(12) of Regulation (EU) 2022/2065 and that the fact that it gradually enhanced its processes does not indicate non-compliance.[442] The provider of X describes the January 2024 update of its processes as encompassing '*a newly designed internal ticket escalation process to significantly improve correspondence traceability, applications assessment, and overall efficiency of the application system; a self-imposed* [redacted] *deadline to review service level agreements ('SLAs'); and two more email templates through which researchers could obtain more tailored responses (not only on rejection grounds)*'.[443]

(389) Third, the provider of X maintains that it responds to applicants in less than [redacted] days for 96.4% of the applications and that it '*repl[ies] to almost all-follow up emails from non-rejected applicants in less than* [redacted] *days*'.[444]

(390) Fourth, the provider of X disagrees with the Commission's assessment that the design and implementation of the 40(12) application form created unnecessary friction for applicants. That provider argues that Article 40(12) of Regulation (EU) 2022/2065 does not mandate any specific mechanism to be implemented to enable researchers to request access, and that that Regulation does not issue any guidance on this specific topic.[445] In addition, that provider argues that it is under no obligation to automatically

439 Reply to the Preliminary Findings, paragraph 296.
440 Reply to the Preliminary Findings, paragraph 269.
441 Reply to the Preliminary Findings, paragraph 301.
442 Reply to the Preliminary Findings, paragraph 298 and 299.
443 Reply to the Preliminary Findings, paragraph 270.
444 Reply to the Preliminary Findings, paragraph 300 and 315.
445 Reply to the Preliminary Findings, paragraph 315.

provide researchers with an acknowledgment of receipt and a copy of their request, contrasting Article 40(12) of Regulation (EU) 2022/2065 with the obligations laid down in Article 16 of that Regulation in relation to the notice and action mechanism. The provider of X also argues that '*researchers have been generally provided with a copy of their applications if they request a copy*' and that it has contacted most applicants within [redacted] days following the submission of applications. [446]

(391) Fifth, in relation to the information made publicly available about the possibility to obtain access to data, the provider of X claims that any comparison with Twitter's prior academic data access program is irrelevant, and that the only benchmark should be whether the current processes comply with Regulation (EU) 2022/2065. The provider of X claims that by taking voluntary past practices as a reference point for enforcement purposes, the Commission could create a disincentive for providers to voluntary exceed their legal obligations. The provider also claims that the Regulation does not oblige data providers to train researchers nor to make available specific resources and argues argued that it was '*working with researchers to explain relevant information, such as which kind of data can be extracted*', which was not considered in the Preliminary Findings.[447]

(392) In the fifth place, the provider of X claims that it provided adequate access quotas and durations to Qualified Researchers after 9 November 2023.

(393) As regards access quotas, the provider of X claims that giving access to 1 000 000 posts per month satisfies the requirement set out pursuant to Article 40(12) of Regulation (EU) 2022/2065. That provider references application [redacted], where applicants requested access to 250 000 posts per month.[448] That provider claims that other providers of very large online platforms impose much stricter quotas on Qualified Researchers.[449] The provider of X also claims that the fact that it provided much more extensive access under the former Twitter Academic API program does not prove that the current quota contravenes Article 40(12) of Regulation (EU) 2022/2065.[450] Lastly, that provider claims that it has considered granting quota extensions when requested by applicants and has proactively enquired whether its quotas were sufficient.[451]

(394) As regards access durations, the provider of X claims that an initial access period of 6 months is proportionate and complies with Article 40(12) of Regulation (EU) 2022/2065, noting that some applicants requested 6 months of access or less. That provider considers that Article 40(8), point (e), of that Regulation places on researchers the burden of proving that the time frames requested are necessary and proportionate for the purposes of their research. In the alleged absence of guidance, that provider argues that access durations must be determined on a case-by-case basis.[452]

(395) The provider of X also claims that even if it has established an initial 6-month duration for data access, it is amenable to extend this period where necessary for and proportionate to the purposes of the research. That provider claims to have '*clearly*

446 Reply to the Preliminary Findings, paragraph 318.
447 Reply to the Preliminary Findings, paragraph 314.
448 Reply to the Preliminary Findings, paragraph 311.
449 Reply to the Preliminary Findings, paragraph 311.
450 Reply to the Preliminary Findings, paragraph 312.
451 Reply to the Preliminary Findings, paragraph 313.
452 Reply to the Preliminary Findings, paragraph 304.

disclosed this information to newly accepted researchers', to '*invite justifications on the need for longer access*',[453] and to have proactively sent follow-up emails to Qualified Researchers to offer extensions to data access when the initial period was close to expiring.[454] The provider of X also claims that it grants provisional access duration extensions to Qualified Researchers while it reviews their extension requests.[455] Lastly, the provider of X argues that it has granted extensions to '*all four researchers that requested additional access, as they continued to meet the requirements in Article 40(12)',* which happened after the Commission issued Preliminary Findings.[456]

(396) In the sixth place, the provider of X claims that the Commission fails to prove that its practices have a dissuasive effect on researchers.[457]

(397) In the seventh place, the provider of X claims that it does not prohibit scraping to Qualified Researchers, has legitimate reasons not to allow indiscriminate scraping, and that Regulation (EU) 2022/2065 does not prevent providers of very large online platforms from prohibiting scraping

(398) First, the provider of X claims that the Commission has misinterpreted the wording of X's Terms of Service, which '*do not impose an absolute ban on scraping*'. That provider insists that section 4 of its terms prohibits '*crawling and scraping'* '*unless you have been specifically allowed to do so in a separate agreement with us',* and that nothing prevents independent researchers to reach such an agreement.[458] The provider of X also notes that at the date of its reply to the Preliminary Findings, it had not received any request from researchers to scrape data that is publicly accessible in X's interface.[459]

(399) Second, the provider of X argues that its prohibition of independent access (such as by means of scraping) complies with Union law and that providers of online platforms have legitimate reasons to prohibit scraping. The provider of X claims that the Preliminary Findings mandate '*indiscriminate access to scrapers*' and that mandating providers of very large online platforms to allow scraping would *'constitute an unjustifiable interference with their right to property and their freedom to conduct their business'.*[460]

(400) According to the provider of X, allowing '*indiscriminate data scraping would significantly harm X, its systems and its ecosystem*', as scraping generates '*millions or even billions of requests, straining the capacity of X's servers and disrupting the experience of legitimate users*', as opposed to access through X's APIs, which are '*optimised for high-volume automated requests*'.[461] That provider also argues that scraping '*can lead to the emergence of fake accounts, spam proliferation, and real-time surveillance threats that compromise user privacy*'. In addition, '*indiscriminate access to scrapers, as mandated by the PF, could also result in lost revenue from*

453 Reply to the Preliminary Findings, paragraph 305.
454 Reply to the Preliminary Findings, paragraph 306.
455 Reply to the Preliminary Findings, paragraph 307.
456 Reply to the Preliminary Findings, paragraph 308.
457 Reply to the Preliminary Findings, paragraph 319.
458 Reply to the Preliminary Findings, paragraph 325.
459 Reply to the Preliminary Findings, paragraph 326.
460 Reply to the Preliminary Findings, paragraph 327.
461 Reply to the Preliminary Findings, paragraph 328 and 329.

developers who might otherwise purchase legitimate API Access'.[462] The provider of X gives examples of such abuse, such as '*unqualified researchers seeking greater data access for research that does not comply with Article 40(12) DSA or that pursues hidden commercial interests, businesses desiring unrestricted API access for direct message solicitations or selling increased follower counts, or governments seeking unrestricted API access to monitor public opinion'.*[463]

(401) The provider of X also argues that '*bot presence, fake profiles and data scrapers damage X's trust with its users, who expect X to enforce its policies on privacy and content authenticity. Bots create privacy and security risks, expose users to spam, degrade user experience, and risk users leaving the platform. This, in turn, threatens X with the loss of content, data, and other benefits from a large, active user base*'. That provider cites the cases of Clearview and MegaFace AI, in which private companies scraped large amounts of social media data for commercial purpose without permission from users and service providers.[464] For the provider of X, '*the most efficient way*' to counteract abuse is through '*contractual relationships with third parties and consistent enforcement of access requirements under Article 40(12)*', which it implements by mandating in its Terms of Service that a permission be obtained by any third party engaging in independent access.[465]

(402) Third, the provider of X argues that Regulation (EU) 2022/2065 does not prevent providers of very large online platforms from prohibiting scraping. The provider of X claims that '*the legal debate in the EU has not questioned platforms' right to prohibit this practice through their terms and conditions; it concerns rather the conditions in which this prohibition becomes legally binding on third parties*', pointing to a legal judgment issued in relation to the aviation sector.[466]

(403) The provider of X contests the Commission's claim that independent access is crucial to enable Qualified Researchers to perform specific types of research on systemic risks, such as auditing recommender systems of very large online platforms, which in the case of X cannot be achieved only by accessing data by means of the X API. In challenging that claim, that provider argues that the Regulation does not contain such an obligation, and that the '*interests at stake are so important as to require a public debate before the imposition of any legal obligations'*. It points to the Commission's call for evidence of the Spring of 2023 in that respect, where some stakeholders pointed out that questions remained as to '*under which conditions and procedures*' scraping may fall under Article 40(12) of Regulation (EU) 2022/2065.[467]

(404) For the provider of X, Article 40(12) of Regulation (EU) 2022/2065 '*does not create a right to scrape data publicly accessible at VLOPs' interfaces*', but only stipulates the conditions under which researchers can access data. According to that provider, that provision establishes a '*controlled and purpose-specific access'*, in stark contrast with '*data scraping practices, which often involve automated and indiscriminate extraction of vast amounts of data*'. For that provider, the position of the Commission contradicts the objective of Regulation (EU) 2022/2065, namely, to foster a safer and more

462 Reply to the Preliminary Findings, paragraph 328.
463 Reply to the Preliminary Findings, paragraph 331.
464 Reply to the Preliminary Findings, paragraph 332.
465 Reply to the Preliminary Findings, paragraph 330.
466 The provider of X cites judgment of 15 January 2015, Ryanair v PR Aviation, C-30/14, EU:C:2015:10.
467 Reply to the Preliminary Findings, paragraph 341.

transparent online environment, since data scraping '*can comprise privacy, intellectual property, image rights and security*'.[468]

(405) The provider of X argues that in previous Union law, '*when data scraping has been authorized, it has been under well-defined and specific conditions*', such as was done by the text and data mining exemptions set out in Articles 2 and 3 of Directive (EU) 2019/790 ('the Copyright Directive'). While that provider notes that the Copyright Directive provides for '*an exception or limitation to certain intellectual property rights*', it argues that '*such an exception has not been established with regards to personal data nor to image rights*'.[469]

(406) The provider also claims that the Commission '*incorrectly interprets recital 98 DSA as a legal basis for a limitless authorization allowing scraping of VLOPs content*', where it merely establishes the conditions under which publicly accessible data can be used for research pursuant to Article 40(12).[470]

(407) First of all, the provider argues that recital 98 of Regulation (EU) 2022/2065 makes clear that access rights are conditional upon the researchers meeting the applicable requirements, meaning that '*preventing X from prohibiting scraping on its general terms and conditions is equivalent to forcing X to grant access to all of its content to everyone indiscriminately*'.[471]

(408) In addition, the provider notes that recital 98 of Regulation (EU) 2022/2065 contains language referring to examples of data that should be provided, namely '*for example, on aggregated interactions with content from public pages, public groups, or public figures*'. For the provider, this reference to aggregated data implies '*a specific type of data handling where individual data points are combined to form a summary for analysis*', thus reducing privacy risks and '*contrasting with the invasive nature of direct data scraping, which typically involves accessing detailed, potentially identifiable data*'.[472]

(409) Finally, the provider of X also notes that recital 98 of Regulation (EU) 2022/2065 also requires providers to '*anonymise or pseudonymise personal data except in those cases that would render impossible the research purpose pursued*'. For the provider of X, this requirement underscores the importance of protecting personal data and minimising risks of re-identification. That provider argues that anonymization and pseudonymization are '*fundamentally incompatible with the typical goals and methods of scraping, where data is often collected in its most raw and identifiable form*', meaning that '*Recital 98 does not endorse unfettered scraping',* but rather *'promotes a controlled, ethical approach to data access that prioritises user privacy and the integrity of personal data*'.[473]

6.5. The Commission's assessment of the provider of X's arguments

(410) For the reasons set out in the following subsection, the Commission concludes that the arguments set out in section 6.4 are unable to call into question the Commission's

468 Reply to the Preliminary Findings, paragraph 335.
469 Reply to the Preliminary Findings, paragraph 336.
470 Reply to the Preliminary Findings, paragraph 337.
471 Reply to the Preliminary Findings, paragraph 338.
472 Reply to the Preliminary Findings, paragraph 339.
473 Reply to the Preliminary Findings, paragraph 340.

findings that the provider of X failed to comply with Article 40(12) of Regulation (EU) 2022/2065 as set out in section 6.3 above.

6.5.1. The claims that the provider strikes a reasonable balance with its legitimate interests and other applicable rules in the absence of Commission delegated acts and guidance

(411) In the first place, the claim that data access under Article 40(12) of Regulation (EU) 2022/2065 requires researchers to fulfil the cumulative requirements laid down in that provision does not contradict the Commission's findings in this regard. Contrary to what the provider of X appears to suggest, the Commission never claimed that the requirements of Article 40(12) of Regulation (EU) 2022/2065 are not cumulative, nor does it point to any paragraph of the Preliminary Findings where this is stated or implied.

(412) In the second place, the provider of X's claim that the legal framework governing Article 40(12) of Regulation (EU) 2022/2065 is incomplete and non-operational in the absence of delegated acts is unfounded. Article 40(12) of Regulation (EU) 2022/2065 is a self-standing provision, which sets out clear, directly applicable obligations. It does not require further specification through guidelines or delegated acts for its application.

(413) In its argumentation, the provider of X conflates Article 40(4) and Article 40(12) of Regulation (EU) 2022/2065. It reasons that, since the access mechanism mandated by Article 40(4) of that Regulation is not yet fully operational, the entire data access framework of Article 40, including that of Article 40(12), is rendered unenforceable. That line of reasoning is unfounded. Article 40(4) of Regulation (EU) 2022/2065 entails the creation of a new vetting procedure, relying on national Digital Services Coordinators, to ensure vetted researchers are granted unprecedented access to the non-public data of very large online platforms and of very large online search engines. Conversely, the obligation laid down in Article 40(12) of Regulation (EU) 2022/2065 concerns granting access to publicly accessible data. It does not necessitate any intermediation by Digital Services Coordinators.

(414) In fact, the Commission observes that while, pursuant to Article 40(13) of Regulation (EU) 2022/2065, the Commission is empowered to issue delegated acts laying down the *technical* conditions under which providers '*are to share data pursuant to paragraphs 1 and 4*' of that provision, that empowerment does not extend to Article 40(12) of that Regulation. This confirms that Articles 40(4) and 40(12) of Regulation (EU) 2022/2065 provide for two distinct legal obligations and two different data access mechanisms, which operate independently from one another.

(415) The Commission has reiterated this position to the provider of X, including in its requests for information addressed to that provider on 12 October 2023 [474] and in requests for information sent to 17 providers of very large online platforms and of very large online search engines on 18 January 2024,[475] as well as in its public announcements of its decisions to initiate proceedings related to suspected non-

[474] DSA.100102, Doc ID 237-2.

[475] The Commission publicly communicated on these 17 requests for information in relation to Article 40(12). See https://digital-strategy.ec.europa.eu/en/news/commission-sends-requests-information-17-very-large-online-platforms-and-search-engines-under.

compliance with Article 40(12) of Regulation (EU) 2022/2065 by the providers of five other very large online platforms.[476]

(416) Given that, for the reasons explained above, the obligation to adopt a delegated act does not extend to obligations covered by Article 40(12) of Regulation (EU) 2022/2065, the provider of X's claims that Article 93 of Regulation (EU) 2022/2065 obliges the Commission to adopt delegated acts pursuant to Article 40(13) by 16 November 2022 are equally ineffective. In any event, even if that obligation were to extend to Article 40(12) of Regulation (EU) 2022/2065 (quod non), the arguments advanced by the provider of X are incorrect for the following reasons.

(417) First, Article 93(2) of Regulation (EU) 2022/2065 provides that: '*This Regulation shall apply from 17 February 2024. However, Article 24(2), (3) and (6), Article 33(3) to (6), Article 37(7), Article 40(13), Article 43 and Sections 4, 5 and 6 of Chapter IV shall apply from 16 November 2022*' (emphasis added). In parallel, Article 40(13) of Regulation (EU) 2022/2065 provides: '*The Commission shall, after consulting the Board, adopt delegated acts supplementing this Regulation by laying down the technical conditions under which providers of very large online platforms or of very large online search engines are to share data pursuant to paragraphs 1 and 4 and the purposes for which the data may be used. […]'*. Article 93 of Regulation (EU) 2022/2065 does not require these delegated acts to be in place from the moment of entry into force of that Regulation. It merely sets the earliest date (i.e., 16 November 2022) at which any delegated act supplementing Regulation (EU) 2022/2065, including those pursuant to Article 40(13) thereof, can be issued.

(418) This is confirmed by Article 87(2) of Regulation (EU) 2022/2065, which provides that the delegation of power referred to in Article 40 '*shall be conferred on the Commission for five years starting from 16 November 2022*'. Furthermore, the Commission notes that any delegated act adopted pursuant to Article 40(13) of Regulation (EU) 2022/2065 may not, in practice, be adopted by 16 November 2022. This is so because the exercise of such empowerment requires, in accordance with Article 40(13) of Regulation (EU) 2022/2065, that the Commission consults the European Board for Digital Services prior to adopting delegated acts. However, pursuant to Article 49(3) of that Regulation, Members States only had to designate their Digital Services Coordinators by 17 February 2024, meaning that the Board for Digital Services could not reasonably be expected to have been in place by 16 November 2022. This further demonstrates the absurdity of the provider of X's interpretation of Article 93 of Regulation (EU) 2022/2065, and clearly disproves the claim that the Commission is in breach of its legislative duties by having failed to issue delegated acts by 16 November 2022.

---

476 In addition to X, these investigations relate to Facebook, Instagram, AliExpress, TikTok and Temu. These proceedings were respectively opened on 30 April 2024, 30 April 2024, 14 March 2024, 19 February 2024 and 31 October 2024.
https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2373;
https://ec.europa.eu/commission/presscorner/detail/en/ip_24_1485;
https://ec.europa.eu/commission/presscorner/detail/en/ip_24_1485;
https://ec.europa.eu/commission/presscorner/detail/en/ip_24_926;
https://ec.europa.eu/commission/presscorner/detail/en/ip_24_926;
https://digital-strategy.ec.europa.eu/en/news/commission-opens-formal-proceedings-against-temu-under-digital-services-act.

(419) Second, as already stated above, Article 40(12) of Regulation (EU) 2022/2065 contains a self-executing and directly applicable obligation, with which the provider of X has had to comply since 28 August 2023. Therefore, the provider of X must comply with that provision, regardless of whether the Commission planned to issue any delegated act or guidance on any topic covered by Article 40 of Regulation (EU) 2022/2065.

(420) In the third place, the provider of X's claim that, absent the necessary Commission guidance, it is striking a reasonable balance between granting access to researchers, and other legitimate interests and applicable rules, is unfounded.

(421) As noted by the provider of X, some of the requirements laid down in Article 40(12) of Regulation (EU) 2022/2065 also apply to vetted researchers, namely points (b), (c), (d) and (e) of Article 40(8) of Regulation (EU) 2022/2065.[477] The Commission observes that those requirements refer to well-established practices in the industry, including with regards to providing access to publicly accessible data to researchers, and their presence can therefore be verified by providers of very large online platforms and of very large online search engines without the adoption of a delegated act or any additional guidance by the Commission.

(422) First, as regards the eligibility requirements laid out in points (b) and (c) of Article 40(8) of Regulation (EU) 2022/2065, namely that the researchers are independent from commercial interests and that their application discloses the funding of the research, they largely pre-date the entry into force of Regulation (EU) 2022/2065 and are widely applied to researchers in practice,[478] such as during peer review processes, or when conducting clinical trials. It is therefore no basis for the provider of X to claim that shortcomings in relation to implementing these requirements, such as those referenced in the Preliminary Findings, result from an absence of guidance.

(423) Second, the requirements laid out in points (d)[479] and (e)[480] of Article 40(12) of Regulation (EU) 2022/2065 also build upon well-established practices upon which the provider of X may rely to comply with its obligations under that provision. The Commission notes, in this respect, that many of the applications rejected by the provider of X included extensive considerations and safeguards related to data security and confidentiality safeguards, data protection, as well as research ethics[481] that are in line with those that were already implemented by researchers and research organisations.

(424) Third, access to the publicly accessible data of online platforms is hardly a novel practice. The provider of X can effectively build upon two decades of experience in this respect. Indeed, the provider of X itself pointed to the report of the European

477 Reply to the Preliminary Findings, paragraph 251.

478 For instance, an expert quoted by the Commission in the Preliminary Findings noted that the information asked by the provider of X regarding their funding and organisation '*go beyond traditional industry norms for researcher vetting for public data on social media*', indicating that the eligibility requirements laid out in points (b) and (c) of Article 40(8) of Regulation (EU) 2022/2065 are far from new. ; Mozilla Foundation, https://foundation.mozilla.org/en/blog/EU-Digital-Services-Act-and-The-State-of-X-Transparency/, accessed on 19 June 2024 (DSA.100102, Doc ID 207-161).

479 Namely, for researchers to demonstrate that they are capable of fulfilling data security, confidentiality and personal data protection requirements.

480 Namely, for researchers to demonstrate that their access to the data is necessary for and proportionate to the purposes of their research, and that they expected results will contribute.

481 See an overview of such measures in columns H and K of the table provided in reply to the fourth RFI, Request 32)b) (DSA.100102, Doc ID 160-5).

Digital Media Observatory's (EDMO) Working Group on Platform-to-Research Data Access, released in May 2022 and in which the provider of Twitter participated,[482] to insist that the sharing of data must take place in a manner ensuring an appropriate level of protection for user data. At the time that report was published, the provider of Twitter was providing access to publicly accessible data to thousands of researchers, demonstrating that giving researchers access to data can be reconciled with a serious and constructive approach to ensure compliance with Regulation (EU) 2016/679. In 2023, the provider of X stopped participating in the work of EDMO in relation to researcher access to data, when it withdrew from the Code of Practice against Disinformation, soon after putting an end to the Twitter Academic API program.

(425) The provider of X refers to specific risk scenarios in which researchers could use the pretext of studying systemic risks to gain access to personal or confidential information and then share this information with malicious actors, such as competing businesses or hostile governments, which could then use it to manipulate X's service or its users.[483] However, past public statements from the provider of X demonstrate that it considered research an effective mitigation measure precisely against risks of online platform manipulation by malicious actors.[484] Additionally, to this day, the provider of X continues to provide access to the X API to customers for a fee. Undeniably, the potential risk factors and malicious actors referenced by the provider of X are also very relevant in the context of the X commercial API, as competing businesses and hostile States would naturally have the financial capacity to pay for commercial access. In addition to these specific risk scenarios, the scope of data shared with customers of the commercial API includes the scope of data that is shared with researchers who successfully undergo X's application process, with the same level of granularity and thus a similar likelihood to contain personal or confidential information.

(426) The fact that the provider of X continues to provide API access to commercial customers to this day suggests that it has a degree of confidence in the effectiveness of its measures to ensure data security and confidentiality, to protect its economic interests against unfair competition resulting from the processing of the data requested, and to prevent infringements of the rights to privacy and personal data protection. Therefore, similar verifications could easily be extended to screen and prevent potential abuse of the access granted to the X API under Article 40(12) of Regulation (EU) 2022/2065.

(427) Should the provider of X reasonably suspect or demonstrate the existence of a risk for the protection of personal data or confidential information emanating from a specific application, including by relying on its existing security processes, it would be able to

---

482 Reply to the Preliminary Findings, footnote 286. This footnote mentions the EDMO report on researcher access to data ; https://edmo.eu/wp-content/uploads/2022/02/Report-of-the-European-Digital-Media-Observatorys-Working-Group-on-Platform-to-Researcher-Data-Access-2022.pdf. This 2022 report contains a letter from Twitter which mentions that '*Since 2006, Twitter has had an open API allowing developers and researchers to use data from the public conversation to study diverse topics such as state-backed efforts to disrupt the public conversation, […]'* . (DSA.100102, Doc ID 379).

483 Reply to the Preliminary Findings, paragraph 258.

484 Reply to the Preliminary Findings, footnote 286. This footnote mentions the EDMO report on researcher access to data; https://edmo.eu/wp-content/uploads/2022/02/Report-of-the-European-Digital-Media-Observatorys-Working-Group-on-Platform-to-Researcher-Data-Access-2022.pdf. (DSA.100102, Doc ID 379).

reject the relevant application for failing to fulfil the eligibility requirements set out in Article 40(12) of Regulation (EU) 2022/2065, which, as the provider of X itself recognises, already accounts for such risks. However, to the Commission's knowledge, the provider of X has not implemented even its existing screening processes to applicants seeking access pursuant to that provision, since it has systematically encouraged unsuccessful applicants instead to '*review the different* [commercial] *API tiers available at developer.x.com*', indicating that *prima facie*, it considered such applicants eligible to access its commercial offerings.[485]

(428) Fourth, only a minor share of the rejections issued by the provider of X explicitly invoke concerns related to the measures proposed by researchers to ensure the protection of privacy, security and confidentiality, including the protection of the provider's trade secrets,[486] as opposed to what the provider of X claims.[487] As a matter of fact, that provider's rejection templates, from which it rarely deviates, do not even include standard scripts explicitly pointing to such concerns.[488] In practice, the large majority of the rejections issued by the provider of X contained no more detail than a generic reference to either (i) a lack of sufficient details in the application (which is the primary reason used by the provider of X to reject applicants after the preliminary screening of applications and was not singled out by the Commission), or that (ii) the proposed use of the data was not solely for performing research that contributes to the understanding of systemic risks in the Union, where the Commission argues that the provider of X had an overly restrictive interpretation of the scope of eligible research.[489]

(429) Fifth, and more generally, the EU legislature has clearly established in Article 40(12) of Regulation (EU) 2022/2065 that research contributing to '*the detection, identification and understanding of systemic risks in the Union pursuant to Article 34(1)*' has an important value for the public interest and is a legitimate reason for accessing and processing publicly accessible data. The EU legislature integrated safeguards in that provision in the form of access requirements referring to privacy, data security and confidentiality safeguards, including for the protection of the provider's trade secrets, as well as to the principles of necessity and proportionality of the access with regards to the research objectives. Article 40(12) of Regulation

---

485 This text is included in the standard email to inform applicants about the denial of their request. See 'Version 2.0' of the research access checklist provided in Annex to the reply to the second RFI, Request 5 of section VII (DSA.100102, Doc ID 210-8).

486 Out of the [redacted] applications assessed by the provider of X as of 8 May 2024, only [redacted] out of [redacted] rejected requests explicitly referred to a failure to comply with privacy and security standards, namely in the cases of applications ID [redacted], [redacted] and [redacted], while only [redacted] out of [redacted] rejected requests explicitly referred to concerns regarding the funding and the commercial interests of the applicants, namely in the cases of applications [redacted], [redacted], and [redacted]. See the table provided in reply to the fourth RFI, Request 32)b) (DSA.100102, Doc ID 160-5). This search can be replicated by filtering column R to only include relevant reasons for rejection.

487 Reply to the Preliminary Findings, paragraph 265.

488 Documents TIUC-1771, 1780, 1798, 1818 provided in reply to 'fourth RFI', request 34 (DSA.100102, Doc ID 160-2; 168).

489 See the table provided in reply to the fourth RFI, Request 32)b) (DSA.100102, Doc ID 160-5); The relevant search can be replicated by filtering column R to include reasons for rejections referring to 'insufficient details' and to 'research outside scope', [redacted] out of [redacted] applications assessed as of 8 May 2024 were rejected on these bases.
The Commission discusses the provider's interpretation of the requirement that data be used solely for performing research on systemic risks in the Union in section 6.5.3.1. As of 8 May 2024, this argument had been used in [redacted] out of [redacted] rejections issued by the provider of X.

(EU) 2022/2065 is unambiguous when it comes to publicly accessible data: where researchers fulfil the applicable conditions and where the data is necessary and proportional to conduct the research contributing to the understanding of systemic risks in the Union, then providers have a legal obligation to give access to publicly accessible data. More practically, where the provider of X claims that '*the aggregation and analysis of large volumes of data can reveal complex patterns or trends containing sensitive information*' which can reveal insights that are of a sensitive nature,[490] the Commission notes that an understanding of such complex patterns is often precisely necessary to shed light on the systemic risks stemming from the functioning and the use made of very large online platforms. In such cases, the first solution is not to automatically prevent researchers from accessing data, but rather to ensure that the access takes place under modalities that minimise risks while enabling Qualified Researchers to contribute to the objective of the legal provision, namely the understanding of systemic risks in the Union.

(430) The provider of X notes that recital 97 of Regulation (EU) 2022/2065 contains language requiring access to data be proportionate and appropriate to protect the rights and legitimate interests of very large online platforms, as well as of any party concerned, such as recipients of the service.[491] The provider of X claims that this part of recital 97 advocates for a strict interpretation of the access requirements, without clearly stating what such a strict application entails in practice. However, the Commission considers that recital 97 in fact stresses the importance and legitimate interest of the objective of enabling research on systemic risks in the Union and it ensures that the consideration of the commercial interests of providers should not lead to a refusal to provide access. Indeed, the same recital indicates that, '*whilst without prejudice to Directive (EU) 2016/943'*, where the provider of X has reasonable concerns about its trade secrets, it should '*ensure appropriate access for researchers, including, where necessary, by taking technical protections'*.[492] Additionally, the Commission notes that recital 97 primarily relates to access to non-public data for vetted researchers pursuant to Article 40(4) of Regulation (EU) 2022/2065. The excerpt relied upon by the provider of X is particularly relevant in relation to the possibility for providers of very large online platforms to request amendments to reasoned requests for access, based on security and confidentiality considerations pursuant to Article 40(5) of that Regulation, which is exclusive to the vetting process established pursuant to Article 40(4) of that Regulation.

(431) Recital 98 of Regulation (EU) 2022/2065, by contrast, which specifically relates to access pursuant to Article 40(12) of that Regulation, clearly explains that '*when data is publicly accessible, […] providers should not prevent researchers meeting the relevant criteria from using the data for research purposes that contribute to the detection, identification and understanding of systemic risks'*. Recital 98 further explains that '*providers should anonymise or pseudonymise personal data except in*

490 Reply to the Preliminary Findings, paragraph 263 and 264.

491 Reply to the Preliminary Findings, paragraph 257.

492 According to recital 97 of Regulation (EU) 2022/2065: '[…] *to ensure that the objective of this Regulation is achieved, consideration of the commercial interests of providers should not lead to a refusal to provide access to data necessary for the specific research objective pursuant to a request under this Regulation. In this regard, whilst without prejudice to Directive (EU) 2016/943 of the European Parliament and of the Council, providers should ensure appropriate access for researchers, including, where necessary, by taking technical protections such as through data vaults.'* (emphasis added).

those cases that would render impossible the research purpose pursued'. This shows that data protection safeguards should be appropriate to the scope of the research and should not be used as a means of restricting access to the necessary data. Lastly, recital 98 also encourages providers '*to cooperate with researchers and provide broader access to data for monitoring societal concerns through voluntary efforts*', which, to the knowledge of the Commission, the provider of X has not done since 28 August 2023.

(432) In sum, the Commission considers that the eligibility requirements applicable to access data pursuant to Article 40(12) of Regulation (EU) 2022/2065 build upon well-established practices, as well as upon more than a decade of providing researchers access to publicly accessible data. Those requirements establish safeguards for privacy, data security and confidentiality, including for the protection of the provider's trade secrets. Had the provider of X been truly concerned with user privacy or the confidentiality of the data requested, it should have explored the possibility to implement safeguards, such as those to which recital 98 of Regulation (EU) 2022/2065 refer, instead of refusing access on grounds that are not explicitly related to privacy and confidentiality. Contrary to what the provider of X implies, the majority of shortcomings identified by the Commission cannot reasonably result from such concerns.

(433) The Commission observes, in this respect, that the provider of X had a readily available API with a proven track record at the time that Regulation (EU) 2022/2065 started to apply to its service, while it also had a long history of providing researchers and customers access to its data. Consequently, the provider of X was perfectly able to comply with Article 40(12) of Regulation (EU) 2022/2065. The provider of X's claim that the alleged absence of guidance from the Commission forced it to strike a balance not only fails to alleviate the responsibility of that provider with regards to the shortcomings identified by the Commission, it also hints at a strategy that aims at escaping its obligation to provide access to researchers pursuant to Article 40(12) of Regulation (EU) 2022/2065 and undermining the independent scrutiny of its service.

6.5.2. *The claims that the provider of X did provide free access to the X API to Qualified Researchers from 28 August 2023 to 9 November 2023*

6.5.2.1. The claim that the provider of X was transparent regarding the alleged possibility to request access via 'EU-questions@X.com'

(434) In the first place, the provider of X's claim that it was transparent regarding the alleged possibility to request access via 'EU-questions@X.com' is factually incorrect. None of the resources referenced by the provider of X made a single reference to the possibility for Qualified Researchers to request access via EU-Questions@x.com.

(435) First, the guidance documents mentioned by the provider of X, other than the two web pages mentioned in the Preliminary Findings,[493] were either related to X's commercial API offering or appear to be remnants of the Twitter Academic API program. The latter were largely left unattended by the provider of X, as explained in the Preliminary Findings. For instance, over the relevant period, the researcher communities referenced by the provider of X were largely comprised of researchers trying to

[493] Namely the 'Do Research' page and the 'Academic Research' page https://developer.x.com/en/use-cases/do-research and 'Academic Research' https://developer.x.com/en/use-cases/do-research/academic-research, accessed on 19 June 2024 (DSA.100102, Doc ID 207-62).

understand why their access had been discontinued and asking whether they could still request access. To the knowledge of the Commission, the provider of X did not reply to any of the numerous enquiries originating from researchers over this period, deviating from former practices during the legacy program.[494]

(436) Second, the provider of X only resumed posting in its online academic public discussion boards (or 'forums') in the summer of 2024,[495] namely after the Preliminary Findings were sent to the provider of X. Over the relevant period, the only actions of the provider of X on its academic forums consisted in force-closing a few forums in which researchers were asking for updates about possibilities to access X's data.[496] Likewise, the various '*tailored get starting guides*' available over the period were either dead links, outdated resources related to the terminated academic program, or refer to general (i.e., not '*tailored*') resources made available for clients of the commercial API.

(437) Third, the juxtaposition of a link to general information concerning Regulation (EU) 2022/2065 with an email address presented as the general venue for all '*additional queries*' under that Regulation falls considerably short of constituting effective communication to the research community regarding the possibility to request access to data pursuant to Article 40(12) of that Regulation via this address.

(438) In light hereof, the Commission maintains its finding that, between 28 August 2023 and 9 November 2023, the alleged possibility to apply via EU-Questions@x.com was never effectively communicated by the provider of X to researchers, thereby depriving them of the chance to benefit from it and rendering it ineffective. This is also confirmed by the provider of X's own admission that it did not receive a single request for data access pursuant to Article 40(12) of Regulation (EU) 2022/2065 between 28 August 2023 and 9 November 2023, whereas it received [redacted] requests in the first two months after the introduction of the 40(12) application form. That admission disproves the provider of X's claims that it did provide free access to researchers during this period and that the reason why it did not receive any requests during the relevant period was because applicants simply '*did not apply using the adequate means*'.[497]

494 See examples of researchers seeking information since the Spring of 2023 in the 'academic research' tab of X's online developer forums. Developers working for Twitter/X regularly replied to researchers on these forums under the Twitter Academic API program but have stopped doing so between the spring of 2023 and the Preliminary Findings, while force-closing some academic forums where researchers asked for updates ; X developers, https://devcommunity.x.com/c/academic-research/62, accessed on 29 May 2024 (DSA.100102, Doc ID 181-1; 182-1; 183-1).

495 From the summer of 2024 onwards, the newly resumed activity of the provider of X on its academic forums consisted in mentioning the possibility to apply via the form launched on 9 November 2023; See https://devcommunity.x.com/c/academic-research/62, accessed on 8 October 2025. These forums had been referenced in the Preliminary Findings, see X developers, https://devcommunity.x.com/c/academic-research/62, accessed on 29 May 2024 (DSA.100102, Doc ID 181-1; 182-1; 183-1).

496 X developers, https://devcommunity.x.com/c/academic-research/62, accessed on 29 May 2024 (DSA.100102, Doc ID 181-1; 182-1; 183-1).

497 Reply to the Preliminary Findings, paragraph 321.

6.5.2.2. The claim that the absence of a defined process to assess researcher applications and give access to data does not imply that a request would not have been handled correctly

(439) In the second place, as regards the provider of X's claim that the absence of a defined process to assess researcher applications and to give access to data does not imply that a request would not have been handled correctly, it is impossible to verify this claim due to the lack of effective communication by the provider of X of the possibility to apply via EU-Questions@x.com.

(440) However, the Commission recalls that the provider of X's initial attempt at establishing a process in November 2023 resulted in a partial draft and did not lead to granting access to a single applicant before 26 January 2024. Furthermore, the process established from November 2023 onwards suffered from significant shortcomings, as explained in section 6.3, casting doubt on the fact that should a researcher ever had requested access via EU-Questions@x.com (which was not shown to have ever happened), it would have indeed been able to access data within the relevant period.

6.5.2.3. The claim that no infringement can be inferred from the fact that the provider of X charged researchers a fee for access through its API

(441) In the third place, the provider of X's claim that the lack of compliance with Article 40(12) of Regulation (EU) 2022/2065 cannot be inferred from the fact that the provider of X charged researchers a fee for access through its API is unfounded. The Commission responds as follows.

(442) First, as clarified in recital 96 of Regulation (EU) 2022/2065, the objective of Article 40 of that Regulation is to empower *'investigations by researchers on the evolution and severity of online systemic risks* [since they] *are particularly important for bridging information asymmetries and establishing a resilient system of risk mitigation, informing providers of online platforms, providers of online search engines, Digital Services Coordinators, other competent authorities, the Commission and the public.'* Accepting the imposition of a cost on a tool that has as its objective the establishment of *resilient system of risk mitigation* would endanger the existence of such a tool. Contrary to the provider of X's claim,[498] the mere absence of the specification that providers must bear the costs that are related to the application of the provision does not, in itself, mean that the provider has the right to introduce such a fee. It is clear, for instance, that Digital Service Coordinators and the Commission would not have to pay a fee to access data pursuant to Article 40(1) of Regulation (EU) 2022/2065 even though the provision does not specify whether a fee would be acceptable. The prohibition of the imposition of a cost on qualified researchers is also confirmed by the wording of Article 40 (12) of that Regulation which mandates that providers of very large online platforms or of very large online search engines *'shall give'* access to the relevant data. The choice of the words *'shall give'*, without any indication of reciprocity and understood in their usual meaning, indicates that the access given to data must be without any imposition of financial, or otherwise, exchange.

(443) Second, the provider of X's claim that '*no infringement of Article 40(12) DSA can be inferred from the fact that X charged for researchers' access through its API*' because *'in fact, payment for access to data for research purposes is so widespread and*

498 Reply to the Preliminary Findings, paragraph 321.

accepted that usually specific funding is granted for such purposes'[499] is irrelevant to the imposition of a fee on researchers accessing data pursuant to Article 40(12) of Regulation (EU) 2022/2065. In the first place, the grant brought forward by the provider of X as a proof of the widespread existence of grants dedicated to finance access to data for research purposes[500] is actually offered to all researchers irrespective of whether their research concerns or contributes to the detection, identification and understanding of systemic risks in the Union under Regulation (EU) 2022/2065 referred to in Article 40(12) of that Regulation. There is no reason to believe that a grant offered by [redacted] that is located in [redacted] and therefore outside of the Union would specifically and exclusively address research relative to systemic risks in the Union. As such, the provider of X has simply proven that such grants exist *in general* and outside of research falling under Article 40(12) of Regulation (EU) 2022/2065.

(444) In addition, even if this grant was also offered to researchers that perform their research within the scope of Article 40(12), it would only reflect that *in practice* researchers have to pay a fee to the provider of X rather than indicate the legality of such practice.

(445) Finally, even if imposing a fee to researchers was deemed an acceptable practice under Article 40(12) of Regulation (EU) 2022/2065 (quod non), the fee that was charged by the provider of X would still be in violation of the provision. The level of the fees charged by the provider of X was disproportionate and so high that they prevented Qualified Researchers from accessing and using publicly accessible data to conduct research on systemic risks in the Union, as confirmed by the testimonies of several researchers[501] and by survey data documenting the cessation or cancellation of over a hundred research projects from April to November 2023 following the termination of the Twitter Academic API program that left researchers with no choice but to pay for commercial API access.[502] It is worth noting that the grant mentioned by the provider of X in its reply to the Preliminary Findings is of [redacted][503] and would only cover 1,5 months of access to the 'Pro' tier of the X API (each month of access costing USD 5 000), which is insufficient for the majority of studies on systemic risks in the Union, including for instance the large majority of the twelve projects which the provider of X considered eligible for access pursuant to Article 40(12) of Regulation

499 Reply to the Preliminary Findings, paragraph 321.

500 Reply to the Preliminary Findings, footnote 378 - [redacted], [redacted] Grant', available at: [redacted]

501 Non-confidential minutes of interviews with researchers, 1 March 2024, 26 February 2024, 22 January 2024, 23 February 2024, 22 January 2024, 16 February 2024, 8 February 2024, 6 February 2024, 18 March 2024 (DSA.100102, Doc ID 116-1; 86-1; 36-1; 78-1; 40-1; 103-1; 133; 179; 122).
See the non-confidential testimonies of researchers, 27 February 2024 (DSA.100102, Doc ID 91-1).

502 The termination of the Twitter Academic API program, leaving researchers with no choice but to pay for commercial access to the X API led to the documented cessation or cancellation of over a hundred research projects from April to November 2023 This is a conservative estimate, based on the non-confidential results of a survey provided by a researcher, 26 February 2024 (DSA.100102, Doc ID 63-1).
See also Reuters, https://www.reuters.com/technology/elon-musks-x-restructuring-curtails-disinformation-research-spurs-legal-fears-2023-11-06/, 6 November 2023, accessed on 11 June 2024 (DSA.100102, Doc ID 184-1).

503 Reply to the Preliminary Findings, footnote 378 - [redacted], "[redacted] Grant", available at [redacted]

(EU) 2022/2065.[504] To this end, if it was indeed permissible to charge a fee to researchers for access granted to them under Article 40(12) of Regulation (EU) 2022/2065 (*quod non*), such a fee cannot be part of a provider's commercial strategy nor go beyond covering the costs associated with accessing the data requested. The provider of X did not provide any specific price-range for researchers but rather required for them to request the relevant data through its commercial API under the same conditions and for the same prices as any other third party. As mentioned by the provider of X in its response, the X API has a commercial interest and represents a source of commercial revenue.[505] Finally, and to this end, the Commission notes that other providers of very large online platforms whose core features involve publicly accessible data give Qualified Researchers, free of charge, access to existing commercial APIs or to new APIs rolled out specifically for Qualified Researchers.[506]

6.5.3. The claims that the provider of X's interpretation regarding the applicable requirements for data access was justified

6.5.3.1. Eligible use of publicly accessible data

(446) In the first place, as regards the provider of X's claim that the Commission has failed to establish that its interpretation of the requirement that the data be used solely for performing research that contributes to the detection, identification and understanding of systemic risks in the Union was overly restrictive, the Commission responds as follows.

(447) First, systemic risks are identified in Article 34(1) of Regulation (EU) 2022/2065 as '*<u>any</u> systemic risks in the Union stemming from the <u>design</u> or <u>functioning</u> of very large online platforms and very large online search engines, or from <u>the use made</u> of these services*' (emphasis added). The same Regulation lists four main categories of systemic risks, identified in Article 34(1) of that Regulation, which providers of very large online platforms and of very large online search engines shall include in their service-specific risk assessments. Based on this definition, in practice, a large array of research fields,[507] types, scopes and methodologies[508] can contribute to the detection, identification and understanding of systemic risks in the Union.

(448) Second, the provider of X's criticism of the content of the technical analysis suffers from the same shortcomings as its initial assessment of the applications, displaying a flawed and overly restrictive interpretation of the scope of eligible research pursuant to Article 40(12) of Regulation (EU) 2022/2065, leading the provider to unduly restrict this access. The Commission's technical analysis includes an in-depth review of 4 abusive rejections issued by the provider of X and lists 8 other examples of rejections

---

504 See the table provided in reply to the fourth RFI, request 32)b). This search can be replicated by filtering column Q) to only visualize researchers approved by the provider, and then by consulting column K) which contains the requested access timeframes (DSA.100102, Doc ID 160-2; 160-5). ; See also paragraph 13 of the 'DSA.100102' Technical Analysis (DSA.100102, Doc ID 257-1).

505 Reply to the Preliminary Findings, paragraph 328.

506 EU-US Trade and Technology Council, Status Report: Mechanisms for Researcher Access to Online Platform Data, 5 April 2024, pp 12-17 (DSA.100102, Doc ID 185-1); See for instance the cases of YouTube, Facebook, Instagram, TikTok, LinkedIn.

507 Including but not limited to political science, media studies, economics, psychology, public health, medicine, computer science, gender studies, sociology, law, ethics, linguistics, education, criminology.

508 This includes both quantitative and qualitative methodologies, large-n as well as small-n studies, comparative studies, descriptive, inductive, or explorative studies. Eligible research projects may for instance focus on specific platforms features, use patterns, specific risk areas, to the extent that they contribute to the understanding of systemic risks in the Union.

that appeared unjustified.[509] These 12 rejected applications explicitly referenced by the Commission as being abusive are to be put in perspective with the total of [redacted] applications that were approved by the provider of X between 28 August 2023 and 8 May 2024.

(449) The technical analysis is particularly relevant because the Commission requested, on 8 May 2024, that the provider of X produce the documents underpinning its internal assessment leading to the denial of the four applications covered in the analysis.[510] All 4 applications addressed in the technical analysis were rejected with the exact same generic motive, namely '*Based on your application, it does not appear that your proposed use of X data is solely for performing research that contributes to the detection, identification and understanding of systemic risks in the EU as described by Art. 34 of the Digital Services Act*', without explanation of how this conclusion was reached in relation to each of the respective applications. More specifically, the Commission assessed the provider of X's internal reasoning for each of these specific rejections, which was not effectively communicated to applicants at the time.[511]

(450) **Application** [redacted] relates to a research project [redacted]. This large-scale project is a recipient of funding under the [redacted] Programme[513] and had previously obtained access to data pursuant to Article 40(12) of Regulation (EU) 2022/2065 from other providers of very large online platforms. According to internal documents of the provider of X, the application was rejected because it '*only appear[ed] to pertain to misinformation, but [did] not seem to focus on misinformation around any systemic risk in the EU*'. In its reply to the Preliminary Findings, the provider of X attempted to provide a new justification for its decision, claiming that '[redacted] *primarily concerns* [redacted]*; for research on* [redacted] *to fall within Article 40(12), the researchers at issue should have established some link between the* [redacted] *under investigation and X's structural and operational systems characteristics, such as its algorithms or amplification mechanisms. Absent any direct link, research focusing on* [redacted] *generally would not meet the criteria for systemic risk analysis established in the DSA, falling out of the scope of Article 40(12)*'.[514]

(451) In this argumentation, the provider of X displays its flawed interpretation of the scope of eligible research, and of what constitute systemic risks in the Union, while failing to bring forth any new argument that is not already dispelled in the technical analysis provided by the Commission. The internal justification suggests that the provider of X did not consider '*misinformation*' (in the words of the provider) campaigns to be a relevant systemic risk in the EU in and of itself (while also failing to take into account the fact that the proposed research also focused on disinformation, defined as being intentional, which is typically the case of narratives spread using [redacted]

509 See the examples included in the 'DSA.100102' Technical Analysis (DSA.100102, Doc ID 257-1).

510 See in particular documents TIUC-1633 to 1741 provided in reply to 'fourth RFI', request 34 (DSA.100102, Doc ID 160-2; 168).

511 As discussed in sections 6.3.2.2.2., 6.5.4.1. and in the 'DSA.100102' Technical Analysis (DSA.100102, Doc ID 257-1).

512 [redacted]

513 This was explicitly mentioned by the applicants to the provider of X.

514 Reply to the Preliminary Findings, paragraph 281.

[redacted] and is clearly referenced in recital 84 of Regulation (EU) 2022/2065 as a systemic risk factor).

(452) Finally, as regards the *post-facto* justification put forward by the provider of X that for the project in question to be eligible, applicants '*should have established some link between* [redacted] *and X's structural and operational systems characteristics, such as its algorithms or amplification mechanisms'* (emphasis added), the Commission once more recalls that pursuant to Article 34(1) of Regulation (EU) 2022/2065, systemic risks are identified as stemming from the design or functioning of very large online platforms and very large online search engines, or from the use made[515] of these services. The provider of X omits to mention that systemic risks are defined as also potentially stemming from the use made of the service, which contradicts the provider's claims that research into systemic risks cannot possibly entail the review of user content. In addition, coordinated manipulations of the relevant services are referenced in Article 34(2) of Regulation (EU) 2022/2065 as well as in its recitals 83, 84 and 104 as a factor of systemic risks. Therefore, the argument of the provider of X is not justified and it demonstrates that provider's flawed and overly restrictive interpretation of the scope of eligible research pursuant to Article 40(12) of Regulation (EU) 2022/2065.

(453) **Application** [redacted] relates to a project [redacted]. According to internal documents of the provider of X, the application was rejected because it '*only tangentially relates to 'civic discourse and electoral processes*''. After becoming aware of the technical analysis, the provider of X again stood by its rejection and elaborated on the rather short reasoning contained in internal documents: '[redacted]. *The applicant's research topic relates to* [redacted] *rather than to structural risks directly arising from X's design or operation. For the study of* [redacted] *to fall within the scope of Article 40(12), it would have been necessary for the applicant to demonstrate that their research focused on whether X's structure or operation leads to* [redacted]*, but the applicant failed to provide any information on this point.'*

(454) As noted by the Commission in the technical analysis,[516] research pertaining to [redacted] is largely acknowledged in the research community as relevant to understand the impact of social media on democratic processes. In addition, contrary to the claim of the provider of X, it was clear from the application that the research project in question did purport to understand whether X's systems contribute to [redacted]. Indeed, where a specific platform service relies on recommender systems to personalize content seen by users, such as what is done by the provider of X, any study aiming to [redacted] will inherently study [redacted]. Regardless of this, as described in recital 452 above, systemic risks do not only stem from the design and functioning of the relevant services, but also from the use made of

515 Reply to the Preliminary Findings, paragraph 276.

516 'DSA.100102' Technical Analysis (DSA.100102, Doc ID 257-1).

these services,[517] meaning that establishing a link between the research project and X's structure or operation was not even necessary to qualify for access.

(455) **Application** [redacted] consisted of a project analysing [redacted]. According to internal documents of the provider of X, the application was rejected because '*this research is out of scope. It doesn't pertain to dissemination of illegal content, but rather* [redacted] *as they describe it. It also doesn't pertain to systemic risk of actual or foreseeable negative effects that* [redacted]. *Instead, the research focuses on misinformation generally,* [redacted]'. The provider of X noted that it still stands by the internal assessment of its Developer Team. After becoming aware of the Commission's technical analysis, the provider of X attempted to rationalize its decision *post-facto* by adding the argument that the application was rejected as the researchers explained that their analysis would also include [redacted].[518]

(456) The original internal assessment of the provider of X which led to the rejection as well as that provider's attempt at finding a justification *post-facto* point to a flawed and restrictive interpretation of the scope of eligible research pursuant to Article 40(12) of Regulation (EU) 2022/2065 and point to a deliberate decision to avoid compliance with the obligations stemming from that provision. The initial argument of the provider of X was already addressed in the Commission's technical analysis,[519] including by showing that the applicants had provided extensive and unambiguous justifications on the relevance of their work for the purposes laid down in Article 40(12) of Regulation (EU) 2022/2065 on several occasions, including in their original application. The researchers clearly linked their research to understanding the systemic risk identified in Article 34(1)(c) of that Regulation, related to civic discourse and electoral processes, noting that [redacted] did concern politics. The researchers notably requested access to the X API to retrieve posts which [redacted]. The researchers also explicitly argued that their research would provide useful data for the provider's future risk assessment reports, noting that they had already presented early findings to the team of X engineers developing [redacted], who '*confirmed the robustness and usefulness'* of the project in writing. The researchers noted that according to Regulation (EU) 2022/2065, [520] when assessing systemic risks, providers of very large online platforms should consider how their content moderation processes affect these systemic risks. Applicants also referenced the ongoing Commission investigations related to Articles 34 and 35 of Regulation (EU) 2022/2065, which clearly cover the provider's assessment of systemic risks related to civic discourse and electoral processes and the effectiveness of [redacted] in relation to mitigating such risks, proving that their research was in scope of Article 40(12) of Regulation (EU)

---

517 In the Reply to the Preliminary Findings, the provider of X distorts the legal text by omitting to mention the systemic risks stemming from the use made of the service. This dispels the point of the provider of X that research into systemic risks cannot possibly entail the review of user content or user behaviour; Reply to the Preliminary Findings, paragraph 276.

518 Reply to the Preliminary Findings, paragraph 281.

519 'DSA.100102' Technical Analysis (DSA.100102, Doc ID 257-1).

520 Notably Article 34(2)(b) of Regulation (EU) 2022/2065.

2022/2065. Importantly, [redacted], meaning that they present a specific risk of being abused by coordinated actors. The study proposed by applicants would have therefore contributed to the understanding of systemic risks and was therefore in scope of Article 40(12) of Regulation (EU) 2022/2065.

(457) The provider's *post-facto* claim that the research was out of scope because it considered [redacted] as part of its research sample is also flawed. Indeed, it is justified for a research project to also consider wider geographical contexts where this is necessary to strengthen the contribution of the study for the understanding of systemic risks in the Union, knowing in this case that [redacted] Lastly, as this claim was only formulated by the provider *post*-facto, the provider of X never informed the applicants that it considered part of the public data requested to be out of scope, which may have enabled applicants to challenge this argument, or to revise their application.

(458) **Application** [redacted] relates to a cross-platform research project tracking [redacted]. The project was carried out by a civil society organisation who is a member of the [redacted] as well as a signatory of [redacted] and [redacted].

(459) In this case, a review of the internal documents of the provider of X showed that no arguments had even been formulated prior to issuing the generic denial template to the applicant. The rejection also included an encouragement for the researcher to re-scope their research project to only focus on [redacted] implying that for the provider of X, research on mis- and disinformation on X does not qualify for access under Article 40(12) of Regulation (EU) 2022/2065 unless it is focused on a relevant event.

(460) Such interpretation of Article 40(12) of Regulation (EU) 2022/2065 is incorrect as the risks to civic discourse, public security and public health resulting from mis- and disinformation are evidently not limited to specific time periods, such as [redacted]. The Commission also stresses that cross-platform, [redacted] studies are extremely valuable for the understanding of systemic risks across the Union, in particular since risk mitigation measures may be less advanced in [redacted], and since threats may vary across member states. In its reply to the Preliminary Findings, the provider of X claimed that '*the research focused on tracking disinformation trends without a direct connection to how X's structural elements could contribute or not to these trends'*.[521] Again, this claim demonstrates a clear misunderstanding of the scope of the access requirement as well as of the concept of systemic risks, which is not limited to the design or functionalities of the relevant services, but also to how these services are used, meaning that the research topic does not need be directly connected to *'X's structural elements*' (in the words of the provider) to be eligible. Regardless, [redacted] necessarily interacts with structural elements of the relevant social media, insofar as the

[521] Reply to the Preliminary Findings, paragraph 281.

recommender systems and moderation policy necessarily affect the dissemination of content.

(461) All the applicants covered in the technical analysis of the Commission had complied in a cooperative spirit with the repeated requests made by the provider of X asking them for supporting documents as part of the application process and had provided the provider of X with evidence that they satisfied the eligibility criteria for data access pursuant to Article 40(12) of Regulation (EU) 2022/2065.[522]

(462) Despite challenging the Commission's assessment and supplementing the initial reasoning of its staff in charge of reviewing applications with entirely new arguments *post facto*, the provider of X admits that the 4 applications '*could, at most, potentially point to a small number of isolated alleged errors in the evaluation of a few applications among the many that X received*'.[523] However, the Commission notes that, while the provider of X tries to minimise its errors in assessing the applications covered by the technical analysis, these errors were actually significant, and pointed to a deeply flawed interpretation of the scope of eligible research pursuant to Article 40(12) of Regulation (EU) 2022/2065.

(463) In the second place, in addition to this in-depth analysis, the Commission's finding of an infringement of Article 40(12) of regulation (EU) 2022/2065 is underpinned by a wide base of evidence, including but not limited to exchanges with more than 20 researchers and experts, including interviews with more than 15 of them, as well as by an extensive review of internal documents of the provider of X and of publicly available evidence, all of which are quoted in section 6.3.2.1.1, in the Commission's technical analysis,[524] and throughout this Decision. In gathering evidence in relation to the specific grievance presented in this section, the Commission has privileged researchers and experts with solid credentials and extensive expertise in studying systemic risks in the Union pursuant to Article 34(1) of Regulation (EU) 2022/2065,[525] who nevertheless saw their applications rejected, often with generic explanations. The Commission has never claimed that the provider of X could not reject applications and the Commission's investigation focused on how the provider of X has failed to provide access to data to the group of legitimate beneficiaries pursuant to Article 40(12) of Regulation (EU) 2022/2065, thus preventing them to conduct research on systemic risks in the Union in relation to the provision of X, and hindering the independent scrutiny of its service. The Commission also notes that it interviewed [redacted] applicants approved by the provider of X as of 8 May 2024,[526] and that approved applications were therefore overrepresented in the sample of researchers it has interviewed compared to the entire pool of applicants.[527] This evidence therefore allows the Commission to conclude on the existence of several shortcomings,

---

522 See the examples included in the 'DSA.100102' Technical Analysis (DSA.100102, Doc ID 257-1).

523 Reply to the Preliminary Findings, paragraph 283.

524 'DSA.100102' Technical Analysis (DSA.100102, Doc ID 257-1).

525 I.e. more than 20 researchers and experts, including 7 academic professors, the technical director of a research laboratory comprising more than 30 researchers, as well as the founder of a widely known tool to access platform data.

526 Namely applicants [redacted] and [redacted]; See the non-confidential minutes of interviews with researchers on 19 March 2024, 28 February 2024 (DSA.100102, Doc ID 174-1 ; 113-1).

527 As of 8 May 2024, the provider of X had received [redacted] applications and had determined an outcome for [redacted] applications. The provider of X stated that it had approved [redacted] of these [redacted] applications (4.7% of the total), while it had rejected [redacted] applications (95.3% of the total). See reply to the fourth RFI, request 32, including the relevant documents provided (DSA.100102; ID 160-2; 160-5; 168).

including the provider of X's restrictive interpretation of the requirement discussed in this section. This specific shortcoming is further confirmed by the clarifications and arguments formulated by the provider of X in its response to the Preliminary Findings, including the provider's attempts to justify some of its rejections post-facto, addressed in recitals 452, 455, 457 and 460 as well as the provider's clarification of its interpretation of the requirement that the data be used '*solely for performing research that contributes to the detection, identification and understanding of systemic risks in the Union*', addressed in recitals 465 to 469.

(464) In the third place, as regards the validity of the survey report published by the European New School and the Weizenbaum Institute, which criticised the provider's interpretation of the scope of eligible research projects as being too narrow,[528] the Commission insists that this report has evidentiary value for the present case. The independent study further corroborates the findings of the Commission. First, regardless of the initial intent of the authors regarding the use of their study for enforcement purposes, the report considers the implementation of Article 40(12) of Regulation (EU) 2022/2065 across all providers of very large online platforms and is therefore a relevant piece of evidence. Second, while the report may have sampling biases and a small sample size, it does not invalidate its specific findings that X appears to often rely on one specific explanation for rejecting applications[529] and that the wording of these rejections is unnecessarily narrow and lacks specific explanations.

(465) In the fourth place, by insisting that the use of data provided pursuant to Article 40(12) of Regulation (EU) 2022/2065 shall directly and exclusively contribute to the understanding of systemic risks in the Union, and that any data usage under the provision needs to demonstrably advance the purpose of the research (emphasis added), the provider of X distorts the meaning of that provision and undermines its objective, which is to enable research that *contributes to the understanding of systemic risks in the Union*.

(466) The Commission notes that the eligibility requirement laid down in Article 40(8)(f) of Regulation (EU) 2022/2065 is not included among the applicable requirements to qualify for access pursuant to Article 40(12) of that Regulation, whereas it is one of the applicable requirements for *vetted* access pursuant to Article 40(4) of that Regulation. In that context, Article 40(8)(f) of Regulation (EU) 2022/2065 strengthens the requirement relating to the use of the data to understand systemic risks in the Union, by extending it to '*planned research activities*' of the applicant.

(467) In contrast to Article 40(8)(f) of Regulation (EU) 2022/2065, Article 40(8)(e) of that Regulation sets out the conditions under which access pursuant to Article 40(12) of that Regulation can be considered necessary and proportionate and requires that the '*expected results of the research will contribute to the purposes* [of detecting,

---

528 Weizenbaum Institute, https://www.weizenbaum-library.de/items/39a490c3-f3c1-42a7-a530-318b17e9de49, April 2024, in particular 'Sample Case 1' on Page 5, and page 6 of the report (DSA.100102, Doc ID 128-1): '*it is an unnecessarily narrow interpretation of the provision to restrict data access by claiming a request is not specific enough without further explanations*'.

529 This was confirmed by the table provided in reply to the fourth RFI, Request 32)b) (DSA.100102, Doc ID 160-5); Indeed, as of 8 May 2024, 123 out of 243 applications for data access rejected by the provider of X were rejected on the basis that '*it does not appear that* [the researchers'] *proposed use of X data is solely for performing research that contributes to the detection, identification and understanding of systemic risks in the EU as described by Article 34 of* [Regulation (EU) 2022/2065]'.

identifying and understanding systemic risks in the Union]' (emphasis added). Therefore, by contrasting these two provisions, only the latter of which is applicable to researchers accessing data pursuant to Article 40(12) of Regulation (EU) 2022/2065, it transpires that the interpretation of the provider of X, namely that applicants should not only demonstrate that their research will contribute to the understanding of systemic risks in the Union, but also demonstrate that any data usage will directly and exclusively advance the purpose of the research, is too restrictive.

(468) Additionally, the Commission notes that the requirement for researchers to make research results publicly available which stems from Article 40(8)(g) of Regulation (EU) 2022/2065 is not included among the applicable requirements to qualify for access pursuant to Article 40(12) of that Regulation. This supports the interpretation that access pursuant to Article 40(12) of Regulation (EU) 2022/2065 should not necessarily be conditioned on the publication of any research, but could also serve for exploratory activities, for instance to help preparing a request for vetted access pursuant to Article 40(4) of that Regulation. However, in practice, the provider regularly asks applicants to reply to questions or provide documents which presuppose that the project be at a relatively advanced stage,[530] such as requesting a research proposal.

(469) The Commission observes that, contrary to what is claimed by the provider of X, the requirement that the data be used '*solely for performing research that contributes to the detection, identification and understanding of systemic risks in the Union*' presupposes the fulfilment of two conditions, namely that: (i) the data are used solely for conducting research (as opposed to any secondary purpose, such as commercial purposes, like training a commercial AI or malicious purposes); and (ii) the research mentioned in point (i) contributes to the understanding of systemic risks in the Union.

6.5.3.2. Geographic location of the researchers

(470) In the second place, as regards the provider of X's claims that it did not systematically reject applications on the sole basis that applicants were established outside the Union, and that it is justified to reject researchers solely because they are established or located outside the Union, the Commission observes the following.

(471) First, in its reply to the Preliminary Findings, the provider of X obfuscates the distinction between the requirement that the research contributes to the understanding of systemic risks in the Union (assessed in section 6.5.3.1), and the requirement that researchers be established in the Union, which is not listed as part of the subset of criteria laid down in Article 40(8), points (b), (c), (d) and (e), of Regulation (EU) 2022/2065, as required in Article 40(12) of that Regulation for Qualified Researchers, nor as part of the remaining conditions listed in Article 40(8) of that Regulation applicable to vetted researchers.

(472) The two research projects referenced in the Preliminary Findings were clearly focused on systemic risks in the Union, and were rejected on the latter basis, which does not exist in the Regulation. Indeed, this is clearly confirmed by the rejection letters sent to these applicants, for instance: '[…] *X will limit API access under Art. 40(12) to*

530 Relevant examples include applications [redacted].; (DSA.100102, Doc ID 160-5; 168).

organizations located in the EU.' [531] Internal data of the provider further confirms that the two applications were not rejected due to a failure to fulfil the '*EU Nexus*' requirement referenced by the provider of X, which would correspond to the claim in the provider's reply to the Preliminary Findings, but rather that applicants were rejected due to a requirement labelled '*Affiliated Organization outside EU*'. [532]

(473) Second, where the provider of X claims that it has approved application [redacted] which allegedly included '*non-EU researchers*' (in that provider's words),[533] it must be noted that all applicants were in reality affiliated with EU universities, unlike what is claimed by the provider of X. Furthermore, in relation to the application referenced by the provider of X, only one of the applicants (an [redacted] national affiliated to a [redacted] university) was effectively granted API access, after having waited 4 weeks following the notification that their application had been accepted.[534] The provider of X refused to grant access to a second researcher (a second [redacted] national affiliated to a [redacted] university) who was part of this application and had also requested access.

(474) In apparent deviation from its previous claims that it did not systematically deny access to researchers established outside the Union, the provider of X also argues that the legal text supports '*X's interpretation*', namely that it is justified to reject researchers solely because they are established or located outside the Union.[535]

(475) The Commission notes in this regard that Article 40(12) of Regulation (EU) 2022/2065 only imposes a geographical requirement in relation to the scope of the research itself, as providers of very large online platforms and of very large online search engines are required to give access to publicly accessible data to Qualified Researchers using the data '*solely for performing research that contributes to the detection, identification and understanding of systemic risks in the Union pursuant to Article 34(1) [of Regulation (EU) 2022/2065*'.

(476) Third, the Commission observes that researchers located outside of the Union or affiliated to an institution located outside of the Union can indeed perform research that contributes to the detection, identification and understanding of systemic risks in the Union within the meaning of Article 34(1) of Regulation (EU) 2022/2065 and are therefore eligible to access publicly accessible data, insofar as they fulfil all the requirements set out in Article 40(8), points (b), (c), (d) and (e), of that Regulation.

(477) Where Qualified Researchers outside the Union request access to personal data and such access includes a transfer of personal data outside the Union, the provider of X must ensure that such transfers comply with Regulation (EU) 2016/679, most notably Chapter V of that regulation. However, the Commission notes that the provider of X

---

531 See in table provided in reply to the fourth RFI, request 32)b) (DSA.100102, Doc ID 160-2; 160-5); Applicant [redacted]'s request was denied on the basis that '*until we receive any contrary legislation or guidance, X will limit API access under Art. 40(12) to organizations located in the EU. Accordingly, your application has been denied*.'.
Applicant [redacted] provided additional information on 17 January 2024 and did not receive a reply until 20 February 2024. His request was also rejected on the sole basis that '*Based on your application, it appears your organization is outside of the European Union*.'.

532 See the table provided in reply to the fourth RFI, request 32)b) (DSA.100102, Doc ID 160-2; 160-5).; in particular column R which contains X's internal labels corresponding to the reasons invoked for denying access.

533 It is unclear whether 'non-EU researcher' designates researchers who are citizens of a non-EU state, or researchers that are established outside of the EU.

534 See Doc TIUC-465 provided in reply to 'fourth RFI', request 34 (DSA.100102, Doc ID 160-2; 168).

535 Reply to the Preliminary Findings, paragraph 289.

made no mention of such concerns when denying applications on the basis of their geographical location.[536] As a matter of fact, the applicants rejected by the provider of X on this basis [redacted].[537] The Commission also notes that the provider of X provides commercial access to the X API to clients established outside of the Union, indicating that it has procedures in place to ensure that international data transfers are carried out in compliance with applicable regulations.

(478) For all these reasons, the provider of X's claim that it can deny applications solely on the basis that they originate from researchers established outside of the Union is clearly in contravention of Article 40(12) of Regulation (EU) 2022/2065. In two of the rare instances where applicants established outside of the Union passed the first screening,[538] the provider of X rejected them on the sole basis of their geographical location despite repeatedly requiring them to provide extensive additional information, whereas it was aware of the country of establishment of researchers since the very beginning of the process. Such practices do not only contradict the objectives of Regulation (EU) 2022/2065 and of its Article 40(12) in particular, but also contribute to the opacity and perceived arbitrariness of the processes set by the provider of X. In its Preliminary Findings, the Commission referred to these cases and clarified that the Regulation did not allow for such a restriction.

6.5.3.3. Affiliation to a research organisation

(479) In the third place, as regards the provider of X's claim that, while its internal documents may point to guidance to reject researchers who are not affiliated with a not-for-profit body, organisation or association, it has not implemented them in practice,[539] the Commission responds as follows.

(480) First, accepting this argument would mean that the Commission must wait for the provider of X to implement a flawed data access policy before enforcing the Regulation, which is absurd.

(481) Second, Article 40(12) of Regulation (EU) 2022/2065 does not distinguish Qualified Researchers based on their institutional affiliation or lack thereof. For that reason, Article 40(12) of Regulation (EU) 2022/2065 prohibits providers of very large online platforms and very large online search engines from denying access to publicly accessible data to Qualified Researchers solely based on their (non-)affiliation. This is particularly relevant not only because of the internal guidelines of the provider discussed above, but also since the application form implemented by the provider of X features two mandatory fields related to the organisation to which the researchers are affiliated, [540] as explained in the Preliminary Findings.

---

536 See footnote 531 above.

537 [redacted]; see applications [redacted] and [redacted] the table provided in reply to the fourth RFI, request 32)b) (DSA.100102, Doc ID 160-2; 160-5), in particular column H summarising the applicable data security and confidentiality safeguards.

538 As of 8 May 2024, only [redacted] out of [redacted] applications had passed the first screening; See in table provided in reply to the fourth RFI, request 32)b) (DSA.100102, Doc ID 160-2; 160-5).

539 Reply to the Preliminary Findings, paragraph 297.

540 Namely, the '*Name of Organization*' and '*Organization Affiliation*' fields.

6.5.3.4. Disclosure of funding and independence from commercial interests

(482) As regards the claim of the provider of X that its application form did not contravene the requirements of Article 40(12) of Regulation (EU) 2022/2065, and that the elements singled out by the Commission, namely the information listed as part of the '*Funding*' and '*Commercial Interest*' fields of X's application form, were not mandatory, the Commission considers that this specific conduct of the provider of X is not relevant for the purpose of establishing an infringement of Article 40(12) of Regulation (EU) 2022/2065.

(483) The Commission notes, however, that the '*Commercial Interest*' field of the 40(12) application form is mandatory, and that the provider's instructions for the field are ambiguous as to the optional nature of providing the detailed information singled out in the Preliminary Findings. [541] In addition, researchers are regularly asked to disclose their funding sources and demonstrate their independence from commercial interests, for instance in conflict-of-interest forms or prior to publishing in an academic journal. However, such processes typically do not require researchers to disclose detailed information about their organisation(s) board, its membership, shareholders or grant recipients, as is done by the provider of X. Furthermore, the expert quoted by the Commission in the Preliminary Findings had extensive experience of the vetting of such requirements, supporting the relevance of their comments regarding fact that some of the information requested by the provider of X appeared unusual.[542]

6.5.4. The claims that the provider of X's process to assess applications and grant Qualified Researchers access to data did not suffer from significant shortcomings after 9 November 2023

6.5.4.1. The claim that the provider provided all applicants with the reasons for which their application was not accepted, and that the Commission has failed to explain why this explanation is insufficient

(484) As regards the provider of X's claim that it provided all applicants with reasons for which their application was not accepted, and that the Commission has failed to explain why those explanations were insufficient, the Commission responds as follows.

(485) First, that claim fails to address the Commission's preliminary finding that the provider of X's systematic practice of providing generic explanations for its denials prevents researchers from understanding the reasons why their applications were denied, to challenge their rejection, and/or to adapt their application, thereby hindering their ability to make use of their right pursuant to Article 40(12) of Regulation (EU) 2022/2065. Indeed, the provider's failure to provide meaningful explanations to applicants aggravates the effect of the shortcoming observed in section 6.3.2.1.1, namely the provider's flawed interpretation of the requirement that researchers use the

---

[541] The instructions for this mandatory field ask applicants to '*provide detailed evidence that substantiates [their] answer. […] 'relevant information includes but is not limited to […] (i) purposes and mission (ii) information about your board members, officers, and individuals with access to X data, (iii) previously published research, (iv) affiliation with other organizations, networks, or causes, and (v) the organization's membership, shareholders or grant recipients*' (emphasis added). See https://docs.google.com/forms/d/e/1FAIpQLSdo0O-D6Kxa3cV4g1JLz2T_0Sk3hdEnTdv8dJmibagCnzJ7kg/viewform, accessed on 19 June 2024 (DSA.100102, Doc ID 207-70).

[542] Mozilla Foundation, https://foundation.mozilla.org/en/blog/EU-Digital-Services-Act-and-The-State-of-X-Transparency/, accessed on 19 June 2024 (DSA.100102, Doc ID 207-161).

data solely for performing research that contributes to the detection, identification and understanding of systemic risks in the Union, insofar as it obfuscates the reasoning of the provider (or lack thereof) and impedes the scrutiny of the [redacted] rejections issued on this basis as of 8 May 2024.[543] For instance, for the four wrongly denied applications analysed in detail by the Commission on the basis of internal documents, the provider of X's assessment displayed critical shortcomings, including a failure to formulate clear arguments internally as discussed in section 6.5.3.1.[544] However, in such cases the applicants received generic denial emails which did not contain more information than a reference to the fact that their proposed use of data was not eligible for access pursuant to Article 40(12) of Regulation (EU) 2022/2065, in line with the horizontal practices of the provider, hindering their ability to challenge the rejections, and to exercise their right to access data.

(486) Second, the interaction described as a '*back-and-forth*' by the provider of X only happens when an applicant passes the first screening and generally consists in the provider asking additional questions without providing any meaningful feedback. When applicants fail to pass the first screening, which was the case in [redacted] out of the [redacted] applications assessed as of 8 May 2024, they only receive a succinct rejection generally containing a generic reference to the failure to fulfil one of the requirements, or to the incompleteness of the application, stating that X will *not* monitor replies, which contradicts the assertion of the provider that there is '*usually*' a back-and-forth with applicants.[545]

(487) Third, in several instances, applicants tried to seek explanations following their rejection by the provider of X or offered to provide further clarity. However, to the knowledge of the Commission, the provider of X did not reply to those requests.[546]

6.5.4.2. The claim that the provider's processes exceeded what is legally required, including prior to January 2024

(488) As regards the provider of X's claim that its processes for granting access to publicly accessible data exceeded what is legally required, including prior to January 2024, that claim is belied by the evidence presented throughout this section and section 6.3, and in particular by the empirical fact that not a single application was approved by the provider of X before 26 January 2024,[547] almost 5 months after Regulation (EU) 2022/2065 became fully applicable to that provider, demonstrating the ineffectiveness of X's internal processes, which at the very least led to undue delays in granting access to Qualified Researchers until January 2024.

---

543 As of 8 May 2024, [redacted] out of [redacted] rejections issued by the provider of X were justified on this basis. ; See the table provided in reply to the fourth RFI, Request 32)b) (DSA.100102, Doc ID 160-5); This search can be reproduced by filtering Column R to only include reasons which include 'Research outside scope'.

544 See also 'DSA.100102' Technical Analysis (DSA.100102, Doc ID 257-1).

545 According to the table provided in reply to the fourth RFI, request 32)b) only [redacted] out of [redacted] rejected applicants received a 'custom' denial (this can be replicated by filtering column R) (DSA.100102, Doc ID 160-2; 160-5). See the specific examples provided in the 'DSA.100102' Technical Analysis (DSA.100102, Doc ID 257-1).

546 See for instance applications [redacted] This search can be replicated by filtering column T of the table provided in reply to the fourth RFI, request 32)b), (DSA.100102, Doc ID 160-2; 160-5).

547 Table provided in reply to the fourth RFI, Request 32)b) (DSA.100102, Doc ID 160-2; 160-5).

6.5.4.3. The claim that the provider has swiftly replied to applicants

(489) As regards the provider of X's claim that it has swiftly replied to applicants, that claim must be contextualised, insofar as its first reply to applicants is usually an acknowledgment of receipt, which the provider has decided not to automatically send upon the submission of applications.[548]

(490) From January 2024 onwards, when applicants failed to pass the first screening by the provider of X, that provider has indeed generally communicated denials to applicants at the same time as it acknowledged receipt of their applications, approximately 7 days after the applications' submission.[549] However, for the [redacted] (out of [redacted]) applicants that passed the provider's first screening and were asked to provide additional information,[550] the first reply constituting a meaningful step forward in the application process, namely the provider's request for additional documents, regularly took longer than seven days, with subsequent interactions regularly taking place over several weeks. For instance, all [redacted] applicants approved on 22 April 2024 had applied more than 2 months prior.[551]

(491) The Commission also recalls the facts laid out in the Preliminary Findings, namely that the provider of X has repeatedly failed to reply to applicants for more than a month, and sometimes only replied when applicants enquired to the provider about the absence of a reply to their previous email.[552] In contrast, the provider of X imposes a two-week deadline to applicants to reply to all its communications, failing which it considers an application withdrawn.

(492) In addition, the [redacted] researchers to whom the provider of X granted API access (the [redacted] successful applicant received data via a file transfer) were all informed of their approval on the same two dates,[553] pointing to the fact that the provider of X artificially delays the processing of at least some applications, a practice that represents an undue delay in providing data to qualified researchers.

---

548 See section 6.5.4.4.

549 According to the data covering the period until 8 May 2024, [redacted] out of [redacted] applications were rejected after the first screening, and most frequently with a very succinct explanation (see the following section).; See in table provided in reply to the fourth RFI, request 32)b) (DSA.100102, Doc ID 160-2; 160-5).

550 As of 8 May 2024, only [redacted] out of [redacted] applications had passed the first screening; See in table provided in reply to the fourth RFI, request 32)b) (DSA.100102, Doc ID 160-2; 160-5).

551 Table provided in reply to the fourth RFI, Request 32)b) (DSA.100102, Doc ID 160-2; 160-5). ; Search can be replicated by filtering columns A (Application date), L (Date – Acknowledgment of Receipt sent to applicants), M (Date – Additional information requests sent to applicants), O (Date – additional information sent by Applicants) and P (Date – Decision on Application).

552 See for instance applications [redacted] (in particular document TIUC-1331); [redacted] (in particular document TIUC-242 and 465); [redacted] (in particular document TIUC-1014); [redacted] (in particular document TIUC-1145); [redacted] (in particular document TIUC-367); [redacted] in the documents provided in reply to the fourth RFI, request 32 (DSA.100102, Doc ID 160-5 ; 168). See JIRA tickets provided in reply to the fourth RFI, Request 34, in particular: document TIUC-1739 – comment written on [redacted] 2024 'Applicant [redacted] will receive an email on [redacted] 2024 informing them that their application is under review.' (DSA.100102, Doc ID 160-5; 168).

553 The [redacted] successful applicants who obtained access to the X API, [redacted] decisions to grant access were communicated to applicants on [redacted] 2024 and 6 were communicated on [redacted] 2024, despite applicants having applied on different dates. The [redacted] researcher who has been granted access saw their request approved on [redacted] 2024, their case differs from the [redacted] others as [redacted]. See the table provided in reply the fourth RFI, Request 32)b) (DSA.100102, Doc ID 160-2; 160-5).

(493) As for the provider's claim that it replies to '*almost all follow-up emails from non-rejected applicants in less than three days*',[554] its wording is at best unclear,[555] and appears to be false regardless of how one is to interpret the sentence, including for the reasons highlighted previously in this section.

(494) More generally, the Commission observes that researchers need visibility and predictability to be able to conduct research and delays in the provision of data may jeopardise the research. Indeed, research has a time-sensitive component, for example in cases where their research concerns risks related to specific events, such as elections or crisis situations. In these specific cases, a delay in the provision of the data hampers and sometimes renders meaningless the performance of such research. Additionally, research, which in practice is often a collaborative and highly coordinated effort, relies on meeting set timelines, both vis-à-vis funding entities and research collaborators.

6.5.4.4. The claim that the provider's application form did not create unnecessary friction

(495) As regards the provider of X's claim that its 40(12) application form does not create unnecessary friction, the Commission considers that this specific conduct of the provider is not relevant for the purpose of establishing an infringement of Article 40(12) of Regulation (EU) 2022/2065.

(496) However, the Commission notes that enabling applicants to obtain a timestamp of the submission of their application, as well as a copy of their application after submitting it, are easily implementable and standard Google Form features.[556] The provider of X nevertheless chose not to implement those features into its application process, despite the fact that some researchers asked for a copy of their application, which, in one case took several days to obtain or sometimes was never even obtainable.[557]

6.5.4.5. The claim that any comparison with the level of information available to researchers under the Twitter Academic API is irrelevant

(497) As regards the provider of X's claim that any comparison with the level of information available to researchers under the Twitter Academic API program is irrelevant, the Commission considers that this specific conduct of the provider is not relevant for the purpose of establishing an infringement of Article 40(12) of Regulation (EU) 2022/2065.

(498) However, the Commission notes that the Twitter Academic API program constitutes an example of good practices, applied by the very provider subject to the investigation. Under the Twitter Academic API program, the provider of X had extensive guidance and resources readily available for researchers, which would have led to minimal costs to maintain in place. The provider of X not only eliminated some resources, but it also

---

554 Reply to the Preliminary Findings, paragraph 300.

555 The Commission is uncertain whether 'follow-up email' means an email containing a question to the provider of if it means any follow-up email exchanged after the initial application.

556 For an overview of the process to enable respondents to a Google Form to obtain a copy of their submission: https://support.google.com/docs/thread/202090575/google-forms-copy-of-responses-not-received?hl=en#:~:text=If%20you%20selected%20When%20requested,not%20get%20a%20response%20receipt and https://www.youtube.com/watch?v=eSAfPZRJ3CU, accessed on 19 June 2024 (DSA.100102, Doc ID 207-77; 207-81).

557 In the example mentioned by the provider, namely that of Application [redacted], the researcher waited 13 days to obtain a copy of their application after requesting it ; Doc TIUC-1646 ; See also the cases of applicants [redacted] and [redacted] who requested a copy and seemingly never obtained it; Docs TIUC-02, TIUC-1259; (DSA.100102, Doc ID 160-5; 168).

maintained dead links in its current pages destined to researchers,[558] meaning that any researcher enquiring about the possibilities to access data would likely find poorly maintained pages crowding out the few resources that the provider of X claims to have made available to applicants.[559]

(499) Exchanging emails with applicants after the submission of an application is not an alternative to maintaining public resources facilitating the formulation and submission of such applications. That is particularly true in the present case, since the provider of X rejects a majority of applications after the initial screening, without offering rejected applicants the possibility to ask for explanations, as explained in section 6.5.4.1.

(500) Most important, and independently of any comparison with past practices, the Commission notes that the degradation of the information made available by the provider of X must be understood in the context of the successive shifts in X's data access practices which have occurred between the spring of 2023 and early 2024.[560] In such a context, providing clear information to the potential beneficiaries of access[561] is particularly important for a provider of very large online platform to fulfil its legal obligation to give access to data pursuant to Article 40(12) of Regulation (EU) 2022/2065.

6.5.5. The claims that the provider of X has provided adequate access quotas and durations to Qualified Researchers after 9 November 2023

(501) The shortcomings discussed in this section concern the [redacted] researchers who qualified for access to X data pursuant to Article 40(12) as of 8 May 2024[562] according to the provider's overly restrictive interpretation of the applicable requirements. Therefore, suppressing the ability of these Qualified Researchers to access necessary and proportionate data volumes and durations is an infringement of the DSA

---

558 'Do Research' https://developer.x.com/en/use-cases/do-research and 'Academic Research' https://developer.x.com/en/use-cases/do-research/academic-research, accessed on 19 June 2024 (DSA.100102, Doc ID 207-62).

559 The lack of usefulness of which has already been discussed in sections 6.3.2.4.1 and 6.5.2.1.

560 For instance, the standard access duration and quotas offered by the provider of X are, substantially inferior to the access that researchers were previously provided under Twitter's academic data access program. To the knowledge of the Commission, these limitations are not conveyed to applicants at any point before or during the application process implemented by the provider of X to grant access to data pursuant to Article 40(12).
Default access quotas are ten times smaller than under the Twitter Academic program, while default access duration is now 6 months as opposed to indefinite or tailored durations under the Twitter Academic program;
See the following archived webpages: The Internet Archive, Twitter API for Academic Research | Products | Twitter Developer Platform, capture of the website as it appeared on 12 February 2023, accessed on 19 June 2024 (DSA.100102, Doc ID 207-141); The Internet Archive, https://web.archive.org/web/20230212182612/https:/developer.twitter.com/en/use-cases/do-research/academic-research/resources, accessed on 19 June 2024 (DSA.100102, Doc ID 207-145).; GitHub maintained by a Twitter/X developer dating from March 2021: GitHub, https://muhark.github.io/python/scraping/tutorial/2021/03/25/getting-started-with-academic-twitter.html, accessed on 19 June 2024 (DSA.100102, Doc ID 207-125).

561 Such as by informing researchers about the standard applicable data access quotas and durations prior to their application, enabling them to justify the need for potential extensions of these standard requirements, which the provider of X did not do, as discussed in section 6.3.2.3.

562 As of 8 May 2024, the provider of X had received [redacted] applications and had determined an outcome for [redacted] applications. The provider of X stated that it had approved [redacted] of these [redacted] applications (4.7% of the total), while it had rejected [redacted] applications (95.3% of the total). See reply to the fourth RFI, request 32, including the relevant documents provided (DSA.100102; ID 160-2; 160-5; 168).

(502) In this respect, the Commission notes that the objective of Article 40(12) of Regulation (EU) 2022/2065 is not the access to data in itself but rather access whose purpose is to perform research that contributes to the detection, identification and understanding of systemic risks in the Union pursuant to Article 34(1) of that Regulation. To this end, it is indispensable that providers of very large online platforms and very large online search engines give access, at the least, to the data that is necessary for the performance of this research where access to such data is proportionate. This is also confirmed by Article 40(8)(e) of Regulation (EU) 2022/2065, which requires the researchers to demonstrate that the data and the time frames requested are '*necessary for, and proportionate to, the purposes of their research*'. The requirement for researchers to be able to provide such an assessment, combined with the objective of Article 40(12) of Regulation (EU) 2022/2065, indicate, indeed, that providers of very large online platforms and very large online search engines must give access to all publicly accessible data that are necessary for, and proportionate to, the purposes of their research.

(503) Therefore, where Qualified Researchers demonstrate to the provider of X that the requested data and timeframes of access are necessary for and proportionate to the purposes of their research, that provider is obliged to grant them the access to the requested data for the timeframes necessary in order to fulfil its obligations under Article 40(12) of Regulation (EU) 2022/2065.

6.5.5.1. Access quotas

(504) As regards the provider of X's claim that it provided adequate access quotas to Qualified Researchers, the Commission responds as follows.

(505) First, as explained in recitals 502 and 503, Article 40(12) of Regulation (EU) 2022/2065 requires providers of very large online platforms and very large online search engines to give access to, at least, the data that are necessary for and proportionate to research projects that respect the conditions set out in the same Article. Quotas are limitations of the amount of data that can be accessed by Qualified Researchers during a given timeframe via an API.[563] As such, where quotas impede the pursuit of research pursuant to Article 40(12) of Regulation (EU) 2022/2065 by not providing all the necessary data for the purpose of a given research project that effectively results in the access to data that are publicly accessible on X's online interface and are necessary for the research's purpose not being granted to researchers in contravention of Article 40(12) of Regulation (EU) 2022/2065.

(506) As explained in recital 393, the provider of X claims that its horizontal imposition of quotas amounting to 1 000 000 posts per month is not contrary to Article 40(12) of Regulation (EU) 2022/2065 because (i) some researchers requested significantly less data than the quotas permit; (ii) providers of other very large online platforms and very large online search engines have more restrictive quotas than X; (iii) the provider of X *'has considered granting quota extensions, when so requested'*; and (iv) the provider of X has '*proactively asked researchers whether the access quota provided was sufficient*'.

(507) These claims are not of a nature to call into question the Commission's finding. Indeed, the existence of research that requires less data than the provider's standard

563 The level of access provided by default to all Qualified Researchers, namely the 'Pro' API access tier, allows researchers to access 1 million tweets per month, as explained in section 6.3.2.3.2.

quota limitations allow for[564] is not an indication that *all* eligible research is possible within the scope of such a quota. In fact, the self-declared willingness of the provider of X to consider '*granting quota extensions, when so requested'* and to '*proactively [ask] researchers whether the access quota provided was sufficient*' indicates that the provider of X is admitting that it is possible that more data are necessary for some research projects carried out on the basis of access granted pursuant to Article 40(12) of Regulation (EU) 2022/2065.

(508) Similarly, while it is true that some providers of other very large online platforms and very large online search engines might have even more restrictive quotas in place, it is also true that each online platform has a specific design and features, which significantly impact user experience, but also affect the different types of research projects that can be conducted across services. For this reason, comparing standard data access quotas applied by other providers of very large online platforms, as done by the provider of X, is of limited relevance. In any event, the provider of X's claims regarding the data access mechanisms of other providers of very large online platforms contains some factual errors.[565] Moreover, some of those providers are currently also under investigation for suspected non-compliance with Article 40(12) of Regulation (EU) 2022/2065 by the Commission.[566]

(509) Finally, the provider of X claims[567] to have considered quota extensions when requested by applicants and to have proactively asked Qualified Researchers whether the access provided was sufficient to conduct their research are factually incorrect. The only example (i.e. application [redacted]) referenced in that regard is an instance where the Qualified Researcher initiated the request for higher quotas themselves.[568] As

---

564 The example of application [redacted] used by the provider was that of a study focusing on [redacted]. As explained in the Preliminary Findings, some applicants clearly explained to the provider that higher quotas were necessary to fulfil their research objectives. In such cases, when the applicants' claims are properly substantiated with regards to the objective of conducting the research, failing to grant sufficient data access quotas is in contravention of Article 40(12).

565 The provider of X seems to refer to the data access programs related to TikTok and to YouTube, which respectively offer access to '*1,000 requests per day across their APIs, allowing up to 100,000 records per day*' for TikTok, and to '*a quota of 10,000 units per day*' for the YouTube API. As regards the TikTok API, the limit is 1,000 requests per day, covering 100,000 records per day, and therefore 3 million of records per month. The provider of X also presents the limit of the YouTube API as being '10,000 units per day', (emphasis added) but fails to mention that the notion of 'unit' represents a number of specific queries to the YouTube API rather than a specific number of pieces of content retrieved via the API. The provider of X also fails to mention that the alleged quotas of the YouTube API that are referenced in its reply to Preliminary Findings correspond to the free tier of access to the YouTube API, which is granted to any applicant, not just to Qualified Researchers. Researchers can request quota increases to the provider of YouTube to exceed the limit referenced by the provider of X.

566 In addition to X, the Commission opened proceedings regarding compliance with Article 40(12) of Regulation (EU) 2022/2065 in relation to Facebook, Instagram, AliExpress, TikTok and Temu. These proceedings were respectively opened on 30 April 2024, 30 April 2024, 14 March 2024, 19 February 2024 and 31 October 2024.
https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2373;
https://ec.europa.eu/commission/presscorner/detail/en/ip_24_1485;
https://ec.europa.eu/commission/presscorner/detail/en/ip_24_1485;
https://ec.europa.eu/commission/presscorner/detail/en/ip_24_926;
https://ec.europa.eu/commission/presscorner/detail/en/ip_24_926.

567 Reply to the Preliminary Findings, paragraph 313.

568 Part of the exchange between applicant [redacted] and the provider of X is available in Annex 1.1 of Reply to the Preliminary Findings, pages 2-10 (DSA.100102, Doc ID 337-1).

explained in the Preliminary Findings, the relevant Qualified Researcher was interviewed by the Commission. According to that researcher, it first reported the issue of insufficient quotas to the provider of X on 3 March 2024 and enquired on whether [redacted]'. That researcher subsequently underwent a series of detailed questions from the provider of X requiring explanations on the necessity of a quota extension. In its response, that researcher reiterated several times that [redacted]. At the request of the provider of X, the researcher then provided [redacted]. However, it seems that, despite the responsiveness of the applicant to the repeated inquiries of the provider of X over several months and according to the correspondence annexed by the provider of X to the Reply to the Preliminary Findings sent on 26 September 2024,[570] the provider of X did not appear to have taken a decision on the extension request at that date, despite the fact that the researcher requested a quota extension on [redacted] 2024.

(510) Second, Article 40(12) of Regulation (EU) 2022/2065 requires that access to data must be given to qualified researchers *without undue delay*. This requirement is not limited to access to *some* data but to all the publicly accessible data that are necessary for and proportionate to the purpose of the research. As such, the undue delays identified in granting quota extensions must be considered as undue delays in giving access to data within the meaning of Article 40(12) Regulation (EU) 2022/2065 and therefore results in the lack of compliance by the provider of X with that provision.

(511) When giving access to publicly accessible data to Qualified Researchers, the provider of X does not assess whether the necessary data would exceed the quota *ab initio,*[571] nor does it do so soon after initial access is granted. As a result, and as explained above, the provider of X relies on individual Qualified Researchers requesting a specific extension of the quota after first obtaining access to the 'Pro' API access tier. This additional and unnecessary procedural step *de facto* adds a procedural and decisional delay to data access pursuant to Article 40(12) of Regulation (EU) 2022/2065 that is not justified by any objective necessity. As such, it unduly delays the process of granting access to data for Qualified Researchers.

---

569 I.e. the standard quotas granted by the provider of X.

570 Part of the exchange between applicant [redacted] and the provider of X is available in Annex 1.1 of Reply to the Preliminary Findings, pages 2-10.

571 During the application review, the provider of X enquires about applicant's data needs. At the end of this process, according to X's internal guidelines as well as based on empirical evidence, the provider always grants successful applicants access to the 'Pro' API tier regardless of the specificities of their application. ; See V2.2 DSA Art 40(12) Research access checklist', page 10 & 18, provided in reply to the fourth RFI, request 33 (DSA.100102, Doc ID 160-2; 168-604).
The provider's regularly takes more than 2 months to grant Qualified Researchers access to the X API following an application for access to data pursuant to Article 40(12) of Regulation (EU) 2022/2065. For instance, all [redacted] applicants approved on [redacted] 2024 had applied more than 2 months prior. See the table provided in reply to the fourth RFI, Request 32)b) (DSA.100102, Doc ID 160-2; 160-5). ; Search can be replicated by filtering columns A (Application date), L (Date – Acknowledgment of Receipt sent to applicants), M (Date – Additional information requests sent to applicants), O (Date – additional information sent by Applicants) and P (Date – Decision on Application).

(512) The existence of this undue delay is further confirmed by the example brought by the provider of X.[572] Indeed, a Qualified Researcher that had requested a quota extension had not received a final reply to its request for over [redacted] months. To the Commission's knowledge, it was the only case of such a quota extension request being lodged and the extreme delay experienced by the Qualified Researcher indicates that the provider of X lacks internal mechanisms allowing the processing of quota extension requests without undue delay irrespective of the whether the applicant would receive the extension or not.

(513) The effect of this undue delay is also compounded by the lack of mention of the existence of the standard quota limitations by the provider of X, as well as the fact that the Twitter Academic API program had 1000% larger quotas than the mechanism that the provider of X has currently put in place to comply with its obligations under Article 40(12) of Regulation (EU) 2022/2065.[573]

6.5.5.2. Access durations

(514) As regards adequacy of the access duration, the Commission responds as follows.

(515) In light of the explanations by the provider of X in its Reply to the Preliminary Findings, the Commission considers that the existing duration of access granted by the provider of X, in and of itself, does not result in the lack of compliance with Article 40 (12) of Regulation (EU) 2022/2065 by that provider.

(516) As the provider of X also argues, the duration of access should be decided on a case-by-case basis in light of the specific needs of the research in question.

6.5.6. The claim that the Commission fails to prove that the practices of the provider discouraged researchers from using their rights pursuant to Article 40(12) of Regulation (EU) 2022/2065

(517) As regards the discouraging effect that the practices of the provider of X have had on Qualified Researchers, the Commission considers that that effect cannot be proven in isolation, since it could also be due to the conduct described in this Decision's remaining sections concerning the provider of X's failures to comply with Article 40(12) of Regulation (EU) 2022/2065.

6.5.7. The claim that the provider of X does not prohibit scraping to Qualified Researchers, has legitimate reasons not to allow indiscriminate scraping and that the Regulation (EU) 2022/2065 does not prevent providers of very large online platforms from prohibiting scraping

(518) As regards the provider of X's claim that it has legitimate reasons not to allow scraping and that Regulation (EU) 2022/2065 does not prevent providers of very large online platforms and of very large online search engines from prohibiting scraping on their services, the Commission responds as follows.

---

572 Part of the exchange between applicant [redacted] and the provider of X is available in Annex 1.1 of Reply to the Preliminary Findings, pages 2-10.

573 Under the 'Pro' API access tier, researchers can access 1 million tweets per month, under the free Twitter Academic API program discontinued in Spring 2023, the limit was at 10 million per month, which means that there has been a ten-fold decrease of this limit. For current X API offerings, see https://developer.x.com/en/docs/twitter-api/getting-started/about-twitter-api, accessed on 19 June 2024 (DSA.100102, Doc ID 207-58). For past API offerings, see The Internet Archive, https://web.archive.org/web/20230212083710/https:/developer.twitter.com/en/products/twitter-api/academic-research, accessed on 19 June 2024 (DSA.100102, Doc ID 207-141).

(519) In the first place, obstacles to independent automated access to data can be separated into two distinct categories. Both *technical* and *contractual* barriers can be erected by providers of online platforms and online search engines to prevent independent automated access to data. Contractual barriers are often included in the terms and conditions, and condition the usage of the online platform or online search engine to the acceptance of a contractual commitment not to independently access data in an automated manner. Such contractual barriers can be found in X's terms and conditions[574] as discussed in section 6.3.3.

(520) In the second place, Article 40(12) of Regulation (EU) 2022/2065 requires providers of very large online platforms and of very large online search engines not only to *give* Qualified Researchers access to the data that is publicly accessible in their online interfaces, for example by means of specific tools or infrastructures, such as APIs or data transfers, but also to refrain from *contractually preventing* Qualified Researchers from independently accessing data available on their public interface, including by automated means (such as data scraping or crawling), and subsequently using the data for the purposes laid down in Article 40(12) of Regulation (EU) 2022/2065. This is clarified by recital 98 of Regulation (EU) 2022/2065, which states that '*where data is publicly accessible, […] providers [of very large online platforms and very large online search engines] should not prevent researchers meeting an appropriate subset of criteria from using this data for research purposes that contribute to the detection, identification and understanding of systemic risks. They should provide access to such researchers including, where technically possible, in real-time, to the publicly accessible data, for example on aggregated interactions with content from public pages, public groups, or public figures, including impression and engagement data such as the number of reactions, shares, comments from recipients of the service.*' (emphasis added).

(521) In this context, independent access includes manual data access techniques, such as manually copying and pasting from the publicly accessible user interface, as well as automated data access techniques such as scraping or crawling. Automated independent access techniques enable researchers to automatically retrieve all types of publicly available data,[575] in real-time and at a granular level, as it appears in X's online interface, without dedicated actions or infrastructure provided for by the provider of X.

(522) As such, a contractual prohibition to independently access data would result in preventing Qualified Researchers from using data that is already publicly accessible and is, therefore, in direct contradiction with Article 40(12) of Regulation (EU) 2022/2065. Even if that was not the case (quod non), automated independent access techniques, such as data scraping and crawling, are indispensable tools for researchers

---

574 For the purposes of this document, 'terms and conditions' mean all clauses, irrespective of their name or form, which govern the contractual relationship between the provider of intermediary services and the recipients of the service, in line with Article 3(u) of Regulation (EU) 2022/2065.

575 For instance, the X API currently does not allow to retrieve all publicly available data (for instance data related to sensitive content interstitials or to the fact that some posts have 'restricted engagement' restrictions, as well as to new features which may not yet be covered by the API). The provider of X admits that '*[t]he X API is designed to replicate all the data publicly available on the platform, but it does not have 100% parity with all public data points. For example, we have not yet completed builds to enable API access to all data related to X Jobs […] or Communities*'. On the other hand, independent access techniques enable to retrieve all data that is publicly accessible.; See the reply to the second RFI, request 8 of section VII (DSA.100102, Doc ID 210-9).

to perform specific types of research on systemic risks in the Union within the meaning of Article 34(1) of Regulation (EU) 2022/2065. Independent automated data access techniques are not comparable with data access through X's API because independent automated access technique uniquely enable Qualified Researchers to access data necessary to study X's recommender and advertising systems and the risks they pose in relation to a variety of systemic risks as defined by Regulation (EU) 2022/2065, notably by probing those systems and observing how they react to specific stimuli, which is instrumental in understanding systemic risks stemming from the design or the functioning of these systems.

(523) In the third place, the provider of X misinterprets the Commission's preliminary findings by considering that the Commission requires an indiscriminate authorisation to scrape the public interfaces of very large online platforms and very large online search engines. The provider of X claims in this respect that not imposing contractual obligations on Qualified Researchers through its terms and conditions would '*allow indiscriminate scraping*', but this is far from what the Commission argued in its Preliminary Findings. Indeed, Article 40(12) of Regulation (EU) 2022/2065 does not mandate an *indiscriminate* access to data but clearly limits it to Qualified Researchers (i.e., only those users who can benefit from Article 40(12) of Regulation (EU) 2022/2065). The imposition of an ex-ante permission in the provider's terms and conditions means that, should these Qualified Researchers independently access data that is publicly accessible in X's online interface, they would be effectively contractually liable towards the provider of X and that in order not to be exposed to such liability they should have requested an ex ante authorisation from that provider even though Article 40(12) of Regulation (EU) 2022/2065 grants them the right to access this data.

(524) The Commission has never denied the provider's right to protect its service against malicious or illegitimate actors, such as commercial and State actors. In fact, a targeted elimination of the prohibition of independent access for Qualified Researchers in X's terms and conditions would not undermine any of the provider's protections against malicious or illegitimate actors. *A fortiori*, such a clarification of X's terms and conditions would precisely differentiate legitimate beneficiaries of access pursuant to Article 40(12) of Regulation (EU) 2022/2065 from third parties which are not eligible for access. For instance, the commercial actors referenced by the provider of X[576] will undoubtedly fail to fulfil the requirements laid out in Article 40(12) of Regulation (EU) 2022/2065. As a result, the number of Qualified Researchers is considerably smaller in size and is different in nature to the numerous third parties mentioned by the provider of X when justifying the necessity for contractual limitations on scraping and crawling. Article 40(12) of Regulation (EU) 2022/2065 does not preclude providers from carrying out technical checks to identify users that are in the process of scraping and are not Qualified Researchers. Therefore, independent automated access for Qualified Researchers would not result in the functional and financial issues that X claims *indiscriminate access* would cause.

(525) The provider of X also claims that independent access could result in *'lost revenue from developers who might otherwise purchase legitimate API Access (as access would not be limited to qualified researchers only)'*.[577] When used to justify the

[576] The provider of X references the cases of Megaface and Clearview; Reply to the Preliminary Findings, paragraph 332.

[577] Reply to the Preliminary Findings, paragraph 328.

exclusion of Qualified Researchers from independent access, and since there is no reason to not limit the contractual exemption to Qualified Researchers, this claim would necessarily mean that the provider of X relies on the contractual prohibition from independent automated access for Qualified Researchers to promote the access granted through its paid API.

(526) In the fourth place, the provider of X asserts that an obligation to not prevent researchers from independently accessing publicly accessible data is contradictory to the obligation to provide '*for example, on aggregated interactions with content from public pages, public groups, or public figures*' (emphasis added) pursuant to recital 98 of Regulation (EU) 2022/2065. This assertion does not hold for two reasons.

(527) First, the fact that the provider of X has an obligation to provide certain data points or even to '*anonymise or pseudonymise personal data except in those cases that would render impossible the research purpose pursued'* (as stated in recital 98 of Regulation (EU) 2022/2065) represents specific obligations linked to its obligation to provide data to researchers, but does not exclude an obligation to '*not prevent researchers meeting an appropriate subset of criteria'* from using the data that is already publicly available, nor from using automated means to retrieve data that is publicly available. In effect, the obligations of providers differ depending on whether they have to act in order to give access to data or they have to simply not prevent data from being used.

(528) Second, the provider of X claims that '*[t]he mention of aggregated data implies a specific type of data handling where individual data points are combined to form a summary for analysis',* but recital 98 in fact clarifies that the elements quoted are *examples* among others of data points that are to be provided and do not represent an exhaustive list of types of data that have to be provided nor can be used to infer a set of criteria that would serve to restrict the data that researchers could access. The provider of X's reasoning contradicts Article 40(12) of Regulation (EU) 2022/2065, which does not impose conditions on the data to be rendered accessible other than the fact that they are publicly accessible on the very large online platform's online interface.

(529) Third, even if it was considered that the providers' obligations were the same when the provider is actively providing publicly accessible data and when the provider is simply not preventing the use of such data (quod non), recital 98 clarifies that anonymization or pseudonymization as means of ensuring data protection must be applied proportionally and should be limited to the extent that they do '*not render impossible the research purpose pursued*.' As explained in the previous recital, independent automated access to data is an indispensable tool for researchers to perform specific types of research on systemic risks in the Union. Consequently, such access is justified pursuant to Article 40(12) of Regulation (EU) 2022/2065.

(530) In the fifth place, the provider of X claims that it would allow Qualified Researchers to scrape its service after they have sought, and obtained, access through a separate agreement with the provider of X. In other words, the provider of X claims that it is amenable to grant exemptions to the contractual prohibition against scraping that it imposes on users in its own terms and conditions.[578] Since an exemption limited

[578] Section 4 of X's Terms of Service stipulate the following: '*You may not do any of the following while accessing or using the Services: […] (iii) access or search or attempt to access or search the Services by any means (automated or otherwise) other than through our currently available, published interfaces that are provided by us (and only pursuant to the applicable terms and conditions), unless*

exclusively to Qualified Researchers can easily be put in place by adjusting X's terms and conditions and since the imposition of an additional step represents an obstacle to access data that necessarily results in a loss of time, the Commission considers this additional step to represent an undue delay to access to data within the meaning of Article 40(12) of Regulation (EU) 2022/2065 in light of the fact that Qualified Researchers can, in practice, independently access publicly accessible data in X's online interface without an intervention of the provider of X (such as by manually collecting data or by automatically scraping such data).[579] In this regard, the imposition of an ex-ante permission in the provider's terms and conditions effectively means that Qualified Researchers independently (and autonomously) accessing data that is publicly accessible in X's online interface may be contractually liable to the provider of X should they not have requested a permission ex-ante. A contractual exemption within the terms and conditions that would specifically be addressed to Qualified Researchers would in practice only apply to researchers that would be entitled to receive this exemption following the provider of X's review of their request for a permission to independently access publicly available data. Arguing the opposite would necessarily mean that the provider of X reserves itself the right to not grant this exemption to researchers that fulfil the conditions set out in Article 40(12) of Regulation (EU) 2022/2065. As a result, conditioning independent automated access to a prior authorisation from the provider of X represents an unnecessary step resulting in the failure to provide access to data *without undue delay* and is consequently non-compliant with Article 40(12) of Regulation (EU) 2022/2065.

(531) In any event, if granting such ad hoc permissions were sufficient to comply with Article 40(12) of Regulation (EU) 2022/2065 (*quod non*), the provider of X would still be in breach of that provision, since its process to assess applications for access pursuant to Article 40(12) of Regulation (EU) 2022/2065 suffers from critical shortcomings, as discussed in section 6.5.3, 6.5.4 and 6.5.5 indicating an inability to adequately and systematically address such requests for permission.

(532) Additionally, the provider of X has not advertised the possibility of requesting such permission nor any process in place for Qualified Researchers to request it. Indeed, X's Terms of Service[580] are ambiguous regarding the possibility for Qualified Researchers to obtain ad hoc authorisations to independently access X's publicly accessible data, including by means of data scraping. The terms containing the general prohibition of independent access, including by automated means, such data scraping

---

you have been specifically allowed to do so in a separate agreement with us (NOTE: crawling or scraping the Services in any form, for any purpose without our prior written consent is expressly prohibited);'

See X's terms of service: https://x.com/en/tos, accessed on 19 June 2024 (DSA.100102, Doc ID 207-46).

579 Conversely, an intervention of the provider of X is operationally necessary to grant Qualified Researchers to the X API, notably with regards to granting them API access credentials.

580 Section 4 of X's Terms of Service stipulate the following: '*You may not do any of the following while accessing or using the Services: […] (iii) access or search or attempt to access or search the Services by any means (automated or otherwise) other than through our currently available, published interfaces that are provided by us (and only pursuant to the applicable terms and conditions), unless you have been specifically allowed to do so in a separate agreement with us (NOTE: crawling or scraping the Services in any form, for any purpose without our prior written consent is expressly prohibited);'*

See X's terms of service: https://x.com/en/tos, accessed on 19 June 2024 (DSA.100102, Doc ID 207-46).

and crawling, do not contain a reference to any process for Qualified Researchers to seek and obtain an exception to this prohibition from the provider of X. In addition, the language added by the provider of X in its developer agreement in relation to data access by Qualified Researchers, and to which researchers must consent when submitting an application to the provider of X via the 40(12) application form, [581] does not appear to create any process for Qualified Researchers to access data independently, as it references '*Licensed Material (including any derivatives thereof) that you retrieve through the X API.'* (emphasis added). The fact that the provider of X has '*not received any request to allow scraping or similar techniques under Article 40(12) DSA*' as confirmed in its Reply to the Preliminary Findings, [582] is an indication that the opacity of its processes actively dissuades researchers from requesting such permission.

6.6. Conclusion

(533) For all the foregoing reasons, the Commission concludes that the provider of X failed to comply with Article 40(12) of Regulation (EU) 2022/2065. Between 28 August 2023 and 9 November 2023, the fees associated with gaining access to data which were publicly accessible on X's online interface under Article 40(12) of Regulation (EU) 2022/2065, in particular in the absence of any alternative for accessing such data, constituted an infringement of Article 40(12) of Regulation (EU) 2022/2065 by the provider of X. The changes implemented by the provider of X after 9 November 2023 did not bring that provider in compliance with the obligation laid down in Article 40(12) of Regulation (EU) 2022/2065, since the processes implemented by that provider to take in and assess researchers' '40(12) applications' suffered from significant shortcomings. The Reply to the Preliminary Findings did not refer to meaningful changes to the practices that infringe on Article 40(12) of Regulation (EU) 2022/2065 by the provider of X. Consequently, the identified infringement persists. Finally, during the entire relevant period, the provider of X has prohibited Qualified Researchers from *independently* accessing X's publicly accessible data, thus failing to comply with Article 40(12) of Regulation (EU) 2022/2065.

---

581 The provider of X clarified to the Commission that *'Researchers who access data by either subscribing to a Developer API access tier or by applying via the Art.40(12) Application form are not bound by this restriction in X's Terms of Service, but instead are bound by the terms in the X Developer Agreement*'. The X Developer Agreement stipulates the following: '***Digital Services Act.*** *Notwithstanding anything to the contrary in this Agreement, to the extent you are provided access to the Licensed Material pursuant to the procedures described in Article 40 of the Digital Services Act (The Digital Services Act) ('DSA'), your access and use of the Licensed Material is limited solely to performing research that contributes to the detection, identification and understanding of systemic risks in the European Union and only to the extent necessary for X to comply with its obligations under the DSA. Any such use of the Licensed Material is non-commercial as described in Section III(B) of this Agreement. You may not disclose, reproduce, license, or otherwise distribute the Licensed Material (including any derivatives thereof) that you retrieve through the X API to any person or entity outside the persons within your organization necessary to perform the research, unless (i) the information is disclosed to the Digital Services Coordinator or other party specifically permitted by the DSA pursuant to the "vetted researcher" status and procedures described in Article 40, or (ii) disclosure is required by law.*'
See the reply to the second RFI, request 3 of section VII (DSA.100102, Doc ID 210-9); Open Terms Archive, https://github.com/OpenTermsArchive/pga-versions/commit/c63b906e573f9f103649430a31a8080d2f7544e3#diff-2d91701cdd46e3d6db2a5cac7bd6531c806a0723a4f01ca8a1a6075eb98b533aR90, accessed on 19 June 2024 (DSA.100102, Doc ID 207-153).

582 Reply to the Preliminary Findings, paragraph 326 (DSA.100101, Doc ID 233; DSA.100102, Doc ID 337-2; DSA.100103, Doc ID 289-2).

7. ALLEGED PROCEDURAL ERRORS IN THE ADOPTION OF THE DECISION

7.1. The allegations of the provider of X

(534) The provider of X argues that the Commission's investigative procedure leading to the adoption of the Decision suffers from severe flaws that undermine its rights of defence. According to that provider, there appears to be a discrepancy between the gravity of the allegations outlined in the Preliminary Findings and the procedural restrictions imposed on the provider of X. According to the provider of X, many of the issues stem from the Commission's own procedural rules, such as those in Implementing Regulation (EU) 2023/1201 and the data room rules, which prioritise administrative convenience over procedural fairness.

(535) The provider of X argues that, despite its repeated objections, these shortcomings have significantly hindered its ability to defend itself, warranting the annulment of any non-compliance decision based on the Preliminary Findings.[583]

(536) The provider of X has grouped its procedural objections into three clusters: (a) access to file issues; (b) incompleteness of the Commission's file and its unreliability; and (c) breach of rights of defence through the Commission's Decision on the Article 5(6) Implementing Regulation Request. These objections are set out in more detail in the subsections below.

7.1.1. Access to file and data room process

(537) According to the provider of X, the Commission has systematically denied it proper access to exculpatory materials in its investigation, severely restricting its ability to build a defence. In that respect, access was only granted to the Specified External Advisers under a highly restrictive data room procedure, which limited their ability to review and analyse evidence. The provider of X's employees were barred from accessing the data room, which was open for only five consecutive working days with rigid entry and exit rules. The process was further hindered by practical issues, such as the unavailability of internet-connected laptops and temporary system crashes.

(538) Additionally, the provider of X argues that the access to the file procedure provided limited time for preparation and forced that provider's Specified External Advisers to juggle multiple procedural requirements in parallel. Moreover, the Commission's restrictions extended beyond access limitations. Legal privilege was disregarded, as the Commission's officials monitored the data room, preventing the provider of X's Specified External Advisers from freely discussing defence strategies. Many documents provided in the data room contained unjustified redactions, and the vast majority of the file did not include confidential information that would justify such restrictions. Despite multiple objections from the provider of X's Specified External Advisers and requests for broader access, the Commission refused to meaningfully justify its approach.

(539) The provider of X cites case law establishing that excessively restricted access to case files constitutes a procedural error that violates the rights of defence. According to the provider of X, the Commission's approach in the present case breaches due process principles by withholding exculpatory evidence and imposing unnecessary procedural barriers.[584]

---

583 Reply to the Preliminary Findings, paragraphs 47-49.

584 Reply to the Preliminary Findings, paragraphs 50-53.

7.1.2. Incompleteness of the Commission's file and its unreliability

(540) The provider of X argues that the Commission is required to compile a complete and accurate case file to ensure a fair and impartial investigation. That provider alleges that, in the present case, the Commission selectively curated its case file to support a finding of infringement, rather than conducting an objective and thorough inquiry.

(541) A key concern of the provider of X is the Commission's reliance on interviews instead of formal requests for information when gathering third-party evidence. Unlike requests for information, which carry legal penalties for providing misleading or incomplete information, interviews afford the Commission significant flexibility to steer discussions and draft post-factum notes that may not fully reflect the interviewees' statements. According to the provider of X, this issue would be exemplified by email exchanges and interview notes where third parties had to push back against the Commission's attempt to frame their input as evidence of an infringement when, in reality, the discussions had a different purpose. Further concerns allegedly arise from the Commission's misleading characterisation of its investigative efforts. Despite claiming to have conducted multiple interviews with former employees, the evidence suggests that only a single interview took place.

(542) Additionally, during the data room procedure, the provider of X's Specified External Advisers would have discovered that key documents were missing from the case file, including confidential versions of non-confidential documents and other relevant evidence. The provider of X contends that this raises serious doubts about whether additional exculpatory materials were omitted, either inadvertently or deliberately. Despite prior assurances from the Commission that the case file contained all collected and created documents, the provider of X argues that these omissions place an undue burden on it to identify gaps in the evidence.[585]

7.1.3. Breach of rights of defence through the Commission's Decision on the Article 5(6) Implementing Regulation Request.

(543) The provider of X claims that the Commission's actions in handling its Article 5(6) Implementing Regulation Request underscore the arbitrary approach to access to the file and a continued overreliance on the restrictive data room procedure. This left the provider of X at a significant procedural disadvantage.[586]

(544) Specifically, the provider of X argues that the vast majority of documents in the Commission's file, including those specifically requested under Article 5(6) of Implementing Regulation (EU) 2023/1201, are not confidential, making the entire data room and Article 5(6) procedures unnecessary burdensome. Rather than granting direct access from the outset, the Commission took six weeks to reject nearly all of the provider of X's requests. The provider of X claims that the Commission's request to prove that each document was indispensable to their defence would contradict due process. The Commission is not in a position to determine what is indispensable for the defence of the provider of X and requiring it to justify its requests at such early stage would be impractical and unlawful.

(545) The provider of X also contends that the Commission's handling of access to documents was inconsistent and arbitrary. In particular, that provider notes that the Commission granted access to 33 documents under Article 5(11) of Implementing

585 Reply to the Preliminary Findings, paragraphs 54-61.

586 Reply to the Preliminary Findings, paragraphs 62-71.

Regulation (EU) 2023/1201, but denied access to 34 other documents covered by the requested submitted to it under Article 5(6) of Implementing Regulation, without providing any explanation as to the reasons for treating these documents differently. The sole document to which access was provided under Article 5(6) of Implementing Regulation was one that should have been disclosed earlier under Article 5(2) of that Regulation, further highlighting procedural failures.

(546) The provider of X also claims that the Commission based its refusals on irrelevant factors, such as that provider's failure to request an extension of the data room duration, even though an extension would not have resolved the underlying issue of restricted access. The fact that the provider of X's Specified External Advisers could summarise some documents did not negate the need for direct access, as reviewing full documents would provide additional insights. Additionally, the Commission arbitrarily denied the request to classify all of the provider of X's external advisers as Specified External Advisers under the data room rules.

(547) Finally, the provider of X submits that the shifting deadlines further compounded the unfairness of the process. The Commission granted additional time to submit the Reply to the Preliminary Findings on 30 August 2024, but only after the provider of X had worked towards a 5 September 2024 deadline for nearly a month, disrupting its ability to plan their Reply to the Preliminary Findings effectively. All these procedural failures would have significantly undermined the provider of X's rights of defence.[587]

7.2. The Commission's response to the provider of X's allegations

(548) For the reasons set out below, the provider of X's claims relating to the access to file procedure followed in this case should be dismissed. The Commission considers that the procedural aspects raised by the provider of X do not amount to procedural errors infringing that provider's rights of defence.

7.2.1. Access to file and data room process

(549) In the first place, contrary to the provider of X's claim that '[the] *highly restrictive terms of access* [to the file] *have prevented the Defendants from properly reviewing and providing their views on the evidence on the Commission's file*',[588] the Commission in fact fully upheld the provider of X's rights of defence and granted full access to the entirety of the file, acting in full compliance with Implementing Regulation (EU) 2023/1201 and with the Terms of Disclosure Decision adopted pursuant to it.

(550) First, upon the provider of X's request of 31 July 2024 to access the file under Article 5(2) of Implementing Regulation (EU) 2023/1201, the Commission promptly gave access to all documents mentioned in the Preliminary Findings the day after the request, *i.e.*, on 1 August 2024. On the same date, the Commission adopted, pursuant to Article 5(3) of Implementing Regulation, the Terms of Disclosure Decision, setting out the terms and conditions for the Specified External Advisers appointed by the provider of X to have access to all documents in the Commission's file, through the set-up of a dedicated data room. Access to the data room was granted on 5 August 2024.

---

587 Reply to the Preliminary Findings, paragraphs 72-79.

588 Reply to the Preliminary Findings, paragraph 51.

(551) Second, the data room procedure was conducted in accordance with Article 5(3)(c) of Implementing Regulation (EU) 2023/1201, which itself balances the objective of safeguarding the right of defence of the provider of the very large online platform or the very large online search engine in question, while respecting third-party rights and ensuring that the Commission meets its obligation to protect confidential information, and those stemming from Article 84 of Regulation (EU) 2022/2065 and Article 339 TFEU in particular. This is also clarified in recital 5 of Implementing Regulation, which states that '*[w]hile the addressee of the Preliminary Findings should always obtain from the Commission the non-confidential versions of all documents mentioned in the Preliminary Findings, the Commission should be able to decide on a case-by-case basis on the appropriate procedure for access to further information in the file. When granting access to the file, the Commission should ensure the protection of business secrets and other confidential information*' (emphasis added).

(552) While the provider of X claims that '*the Commission has consistently refused to provide the Defendants with access to exculpatory material on its investigation file and has only allowed certain specified external advisers of the Defendants access to much of this material on a highly restrictive basis that necessarily impedes the Defendants' ability to effectively exercise their defence rights*',[589] the manner in which the Commission provided access to the file in the present case, including by providing access to the entirety of the file only to specified external advisers was consistent with the regulatory framework set out in Implementing Regulation (EU) 2023/1201. In fact, Article 5(3)(c) of Implementing Regulation provides that *'[p]ersons listed as specified external legal and economic counsel and technical experts shall not, at the date of the Commission decision setting out the terms of disclosure, be in an employment relationship with the addressee or in a situation comparable to that of an employee of the addressee.*' Indeed, Article 5(3) of Implementing Regulation guarantees that the addressee of the Preliminary Findings has access to the investigation file via its specified external advisers, while direct access by the addressee of the Preliminary Findings is foreseen in Article 5(2) of Implementing Regulation and pertains all documents explicitly cited in the Preliminary Findings. Article 5(3) of Implementing Regulation provides for the possibility for the addressee's specified external advisers to access all other documents, including confidential versions of documents which might have been redacted pursuant to Article 6 of that Regulation, thereby enabling the review of all evidence underpinning the investigatory file. These advisers, who are bound by strict confidentiality obligations pursuant to the Terms of the Disclosure Decision, can review the full case file, while ensuring the protection of sensitive business information, and third-party confidentiality rights. Moreover, in exceptional circumstances and upon a reasoned request, the Commission may pursuant to Article 5(6) of the Implementing Regulation grant access to a non-confidential version of any document on the Commission's file not already provided to the addressee under Article 5(2) of the Implementing Regulation, with a view to making such non-confidential version available to the addressee, provided such access is indispensable for the proper exercise of the addressee's right to be heard.

(553) Based on the above, the Commission considers that the provider of X, alongside their specified external advisers, were not only aware of the procedural framework based on Implementing Regulation (EU) 2023/1201 – which was adopted on 21 June 2023 and

589 Reply to the Preliminary Findings, paragraph 50.

published the day after in the Official Journal of the European Union[590] – but they were also informed in advance about the access-to-file process, including the data room setup.[591] In fact, a draft version of the Data Room Rules and of the acknowledgement of the Terms of Disclosure Decision was shared with the provider of X already on 16 July 2024, ensuring transparency and allowing for the necessary preparations.

(554) In the second place, the claim that the data room was made available only for five working days is also misleading. The Specified External Advisers had the opportunity to request an extension before the closure of the data room, but they decided not to do so. Similarly, the rule that the Specified External Advisers could only re-enter the data room during the next session does not constitute an undue limitation but a safeguard meant to prevent unsupervised information flow outside that room, thereby ensuring compliance with the duty of professional secrecy and the ensuing protection of confidential information

(555) In the third place, the provider of X's claim that several hundred documents were only made available for the first time in the data room misrepresents the fact that only documents not mentioned in the Preliminary Findings were subject to this process. In accordance with Article 5(2) of Implementing Regulation (EU) 2023/1201, all the documents mentioned in the Preliminary Findings had already been provided in the first step of the access to file process and they could have been made available by the Commission directly to the provider of X even earlier than 1 August 2024, if only that provider had requested this.

(556) In accordance with Article 5(3) of Implementing Regulation (EU) 2023/1201, all the documents gathered by the Commission until the adoption of the Preliminary Findings in relation to cases DSA.100101, DSA.100102, and DSA.100103 (*i.e.*, in addition to those mentioned in the Preliminary Findings, which had already been provided to the provider of X) were made available for the first time in the data room. Should the provider of X have wished to access these documents earlier, it was open to that provider to submit a request for access earlier. In that connection, in accordance with Article 5(1) of Implementing Regulation, the Commission shall grant access to the file to the addressee of the Preliminary Findings '*upon request*'. The appointment of specified external advisers is only a prerequisite for accessing all the documents in the Commission's file under Article 5(3) of the Implementing Regulation, *i.e.*, not only the documents mentioned in the Preliminary Findings – access to which is governed by Article 5(2) of the Implementing Regulation, but all those – confidential and, without prejudice to Article 5(4) of Implementing Regulation, non-confidential – included in the Commission's investigation file.

(557) The provider of X only specifically requested access to the file – both to documents cited in the Preliminary Findings in accordance with Article 5(2) of the Implementing Regulation (EU) 2023/1201 and to the whole file pursuant to Article 5(3) of that Regulation – for the first time after that provider had appointed its external legal advisers, i.e. eight days after the Commission had notified the provider of X of the Preliminary Findings and invited it to request access to the file. The provider of X therefore had the time it considered necessary to prepare its request for access to the file. As soon as a request pursuant to Article 5(2) of the Implementing Regulation

590 OJ L 159, 22.6.2023, pages 51-59.

591 DSA.100101, Doc ID 204; DSA.100102, Doc ID 267; DSA.100103, Doc ID 231.

(EU) 2023/1201 was submitted to the Commission, on 31 July 2024, the Commission granted access to the documents cited in the Preliminary Findings on 1 August 2024 and this was the first step of the access to file in accordance with Article 5(2) of Implementing Regulation. As mentioned above, the provider of X could have already requested direct access to the documents cited in the Preliminary Findings under Article 5(2) of the Implementing Regulation immediately after their notification, as granting such access by the Commission does not require the appointment of specified external advisers.

(558) In the fourth place, the provider of X's claim that '*The Data Room Report had to be submitted at an early stage in the Defendants' preparation of their observations on the PF, when necessarily the Defendants and their advisers were still assessing the PF and the supporting evidence and when their views on how to defend themselves were still developing*'[592] is without merit. The provider of X's Specified External Advisers had ample opportunity to ensure that all relevant information was included in the Data Room Report before the data room's closure, as *per* the Terms of Disclosure Decision. Moreover, the necessity to prepare the Article 5(6) Implementing Regulation Request concurrently with the data room procedure was a function of established rules under Article 5(6) Implementing Regulation (EU) 2023/1201, which the Commission duly followed.

(559) In the fifth place, the provider of X's contention that there was no possibility for the Specified External Advisers to review confidential versions of the Data Room Report or documents at a later stage is also unfounded. The conditions of access were explicitly outlined in the Data Room Rules which specifically stipulate that '*The Data Room Report must only contain non-confidential information relevant for the exercise of the Addressee's right to be heard in the context of cases DSA.100101 – DSA.100102 – DSA.100103 – X (formerly Twitter). Data, information or documents from the Data Room shall only be included in the Data Room Report in a summarised, paraphrased and aggregated form not allowing conclusions as to the identity of the information provider, even if such data information or documents do not contain business secrets or other confidential information.*'[593] The responsibility was on the Specified External Advisers to ensure that all necessary information was captured before the data room closed or, alternatively, they could have requested an extension of the duration of the data room.

(560) The data room is designed to include all documents not mentioned in the Preliminary Findings and not only documents which can be classified as confidential. This approach serves to streamline and expedite the process by ensuring that the addressee of the Preliminary Findings has direct access to the non-confidential versions of documents relied upon in the Commission's assessment outlined in the Preliminary Findings, while all the other documents – both confidential and non-confidential – are made available to the specified external advisers in a swiftly manner but under strict confidentiality obligations. This framework allows granting the external advisers broader access to the case file than would typically be available under other EU procedural frameworks, such as those governing antitrust[594] and merger control.[595] The

592 Reply to the Preliminary Findings, paragraph 51, sixth bullet point.

593 Data Room Rules, paragraph 14.

594 Pursuant to Article 15 of Regulation (EC) No 773/2004 relating to the conduct of proceedings by the Commission pursuant to Articles 81 and 82 of the EC Treaty, the right of access to the file shall not extend to business secrets, and other confidential information. Therefore, access to the file is limited to

rationale behind this system is also reflected in recital 4 of Implementing Regulation (EU) 2023/1201, which emphasizes the need to reconcile the efficiency and effectiveness of the proceedings, on one hand, and the possibility to exercise the right to be heard, on the other.[596]

(561) In the sixth place, the provider of X's claim regarding redactions to certain documents in the file is also misleading. Out of approximately 3,800 documents made available in the data room, only five contained limited redactions. The Commission also provided the Specified External Advisers with the opportunity to review printed unredacted versions upon request, which they declined.

(562) In the seventh place, as regards the provider of X's claim that '*[t]he Commission made available in the data room only four laptops, which were not connected to the internet, to the EC's network or to each other. Two of the four laptops temporarily crashed during the first two days of the data room procedure*',[597] the limited technical issues were resolved almost instantly,[598] causing no significant disruption to the process.

---

non-confidential information, while confidential information cannot be accessed by either the addressee of a statement of objections or its external advisers (not even within the safeguarded boundaries of a data room).

595 Pursuant to Article 17 of Regulation (EU) No 2776/2024 implementing Council Regulation (EC) No 139/2004 on the control of concentrations between undertakings, the right of access to the file shall not extend to confidential information.

596 In this context, the Commission refers to the Order in Commission v. Amazon (Case C-639/23 P(R), EU:C:2024:277), in which the Vice-President of the Court of Justice held that '*it must be emphasised that Regulation 2022/2065 is a central element of the policy developed by the EU legislature in the digital sector. In the context of that policy, that regulation pursues objectives of great importance, since it seeks, as is apparent from recital 155 thereof, to contribute to the proper functioning of the internal market and to ensure a safe, predictable and trusted online environment in which the fundamental rights enshrined in the Charter are duly protected.*' (para. 155). The Vice-President also underlined the need for speedy application of the Regulation: '*not applying certain obligations laid down by that regulation will lead to a delay, potentially for several years, in the full achievement of those objectives. Not applying those obligations will therefore give rise to a risk of potentially allowing an online environment which threatens the fundamental rights provided for in the Charter to persist and develop*.' (para. 157). These considerations were also emphasized in subsequent Orders of the President of the General Court in Commission v. Aylo Freesites LTD (Case T-138/24 R, EU:T:2024:431, paragraphs 113 and 115) and in in Commission v. WebGroup Czech Republic (Case T-139/24 R, EU:T:2024:475, paragraphs 126 and 128). Therefore, taking this assessment into account, it is also necessary that the process of granting access to the file is conducted as efficiently as possible, which is done, inter alia, by granting the addressee of the Preliminary Findings access to documents that have not been referenced in the Preliminary Findings, i.e., both confidential and non-confidential documents, through its external advisers in a data room.

597 Reply to the Preliminary Findings, paragraph 51, tenth bullet point.

598 The Commission notes that, in both separate instances, its IT support entered the data room two minutes after the technical incidents occurred and verified that the laptops were correctly working again. The Commission stresses that the laptops provided to the Specified External Advisers were brand-new, purchased specifically for that occasion, and they were tested before the start of the data room. In the first instance of laptop-crash, the following day, in the presence of the Specified External Advisers, the Commission's IT support ran a check of the laptop that briefly crashed and confirmed it was an issue related to the memory hardware. The agreement with the Specified External Advisers was that they would continue using the laptop, with more frequent backups and savings of their work. In the second instance of laptop-crash, the Commission's IT support was unable to identify the exact issue, but the Specified External Advisers indicated that they had not lost any work. In both instances, the Commission indicated to the Specified External Advisers that, if the issue had persisted with the same laptops, it would have provided swiftly new laptops. This is documented in the Commission's own reporting about the data room (DSA.100101, Doc ID 284, DSA.100102, Doc ID 377, DSA.100103, Doc ID 324).

Moreover, the availability of four laptops is standard, as was the absence of an internet connection.[599] In accordance with Implementing Regulation (EU) 2023/1201, the primary purpose of the data room is to allow for the review of the Commission's case file in its entirety, while ensuring confidentiality and preventing external communications. These conditions align with standard practice and do not constitute any irregularity or unfair restriction. Similarly, the prohibition on the Specified External Advisers to discuss data room documents outside of the data room and the Commission's right to inspect materials created during the procedure are safeguards to ensure compliance with confidentiality requirements.[600]

(563) In the eighth place, as regards the provider of X's claim that '*[t]he data room procedure did not respect legal privilege. First, the specified external advisers (all lawyers) attending the data room were prevented from discussing the documents made available to them in the data room. Second, Commission officials were constantly present in the data room making it impossible to freely discuss legal strategy or defence arguments*',[601] the Commission observes that the Specified External Advisers had the option to request a breakout room for discussion, and there is no rule in the Terms of Disclosure Decision nor in the Implementing Regulation prohibiting such requests. However, the provider of X's Specified External Advisers did not make any such request. Additionally, within the data room framework, it is essential to ensure that the specified external advisers do not remove any notes or documents containing information reviewed in the data room. This justifies the Commission's right to inspect documents, notes, and other materials. This does not violate legal professional privilege, as erroneously claimed by the provider of X, since the Commission cannot rely on any of these materials in its investigation.

(564) In the ninth place, as regards the provider of X's claim that '*according to the case-law, overreliance on the data room procedure that leads to 'excessively restricted access to the file' is a procedural error that breaches the rights of defence of the defendant(s)*',[602] the Commission observes that that case-law pertains to competition law proceedings under Article 102 of the TFEU, and the reliance on the data room in the case which gave rise to the judgment to which the provider of X refers[603] was done on an a case-by-case basis without there being any specific legal act providing for the procedure of negotiated access to the file, as the one set out in Implementing Regulation (EU) 2023/1201.[604]

---

599 For instance, in the Standard Data Room Rules made available on the Commission's website in the context of antitrust and merger control proceedings, it is indicated that '*PC workstation[s] [is/are] made available […] No external storage medium will be available […] and a local network connecting the PC workstations and the printer will also be made available.*' No connection to the internet is mentioned. See: https://competition-policy.ec.europa.eu/document/download/f94afeca-0fa8-49e1-8a11-59cbb9dc136a_en?filename=data_room_rules_en.pdf (last accessed on 5 February 2025).

600 For instance, in the Standard Data Room Rules made available on the Commission's website in the context of antitrust and merger control proceedings, it is indicated that '*All printouts and notes may be reviewed by DG Competition's officials at any time*'. See: https://competition-policy.ec.europa.eu/document/download/f94afeca-0fa8-49e1-8a11-59cbb9dc136a_en?filename=data_room_rules_en.pdf (last accessed on 5 February 2025).

601 Reply to the Preliminary Findings, paragraph 51, twelfth bullet point.

602 Reply to the Preliminary Findings, paragraph 52.

603 Case T-136/19, *Bulgarian Energy Holding v Commission*, paragraph 1216.

604 In any event, the case-law referred to by the provider of X is currently under appeal by the Commission (Case C – 14/14 P Commission v Bulgarian Energy Holding and Others).

(565) In sum, the Commission's approach in the present case ensured that the provider of X had access to all relevant documents to prepare its defence, while protecting sensitive business information and third-party confidentiality rights and ensuring the efficiency and effectiveness of the proceedings.

7.2.2. Completeness and reliability of the Commission's file

(566) The provider of X's claim that the Commission's file is incomplete and unreliable, due to the lack of requests for information sent to third parties, overlooks the Commission's investigatory discretion under Article 68 of Regulation (EU) 2022/2065. The Commission indeed has discretion to decide whether to use interviews, requests for information, and/or inspections, as appropriate to the investigation before it. In the present case, the Commission primarily relied on gathering its own evidence, as the alleged infringements were self-explanatory.

(567) The provider of X's claim that the minutes of an interview with CheckFirst and the Mozilla Foundation[605] underwent multiple revisions before finalisation and that the Commission influenced their content to strengthen its case is unsubstantiated. Moreover, as explained in recital 244 above, this claim has no bearing on the evidentiary value of those minutes. The draft minutes of the interview are provided because Article 5(3) of Implementing Regulation (EU) 2023/1201 obliges the Commission to grant access to the entire file, which – in the present data room – included drafts and intermediary versions of documents. This further underscores the high level of transparency throughout the process, fully upholding the provider of X's rights of defence.

(568) The provider of X's claim that certain documents were omitted from the file, potentially including exculpatory evidence, is misleading. Out of approximately 3 800 documents, only around 15 were added to the file during the data room process. These additions primarily consisted of draft versions of minutes and a confidential research report. The Commission added these documents promptly, upon identification of the documents by the provider of X's Specified External Advisers, and, in any event, could not and did not rely on them in the Preliminary Findings or its investigation, as they were either drafts or confidential documents. The Commission's investigation is strictly based on final, approved, non-confidential documents.

7.2.3. The Commission's Response to the Article 5(6) Implementing Regulation Request

(569) On the last day of the data room, the provider of X's Specified External Advisers submitted the Article 5(6) Implementing Regulation Request, whereby they requested: (i) further access to non-confidential versions of documents on the Commission's file (in addition to those cited in the Preliminary Findings and already made directly available under Article 5(2) of Implementing Regulation (EU) 2023/1201 with a view to making such documents available to the provider of X as well as the Specified External Advisers outside the data room procedure; and, as a subsidiary request (ii) that all of the provider of X's external counsel be considered Specified External Advisers even if they have not actually accessed the data room or have not been identified as Specified External Advisers. The Commission rejected those requests because they were excessive and inconsistent with the procedural safeguards laid down in Implementing Regulation (EU) 2023/1201.

605 Non-confidential minutes of interview with representatives from CheckFirst and Mozilla Foundation, [redacted] 2024, (DSA.100103, Doc ID 166-1).

(570) In the first place, the Commission handled the provider of X's Article 5(6) Implementing Regulation Request diligently and in accordance with the established legal framework. The provider of X criticizes the Commission for taking too long to provide a response on that request. However, Article 5(6) of Implementing Regulation (EU) 2023/1201provides that '*additional access* [to non-confidential versions of documents made available for the first time in the data room and not under Article 5(2) of Implementing Regulation (EU) 2023/1201] *may only be granted exceptionally and provided that it is indispensable for the proper exercise of the addressee's right to be heard*'. Given that the concept of 'indispensability' was being assessed for the first time in the context of the enforcement of Regulation (EU) 2022/2065, a careful review of the provider's request was necessary. While the provider of X argues that all documents requested under Article 5(6) of Implementing Regulation were non-confidential and access should have been granted, doing so 'in bulk' would have circumvented the procedural framework set up by Article 5(3) of that Regulation, which requires that such documents be made available through the data room process rather than direct access.

(571) In the second place, contrary to the provider of X's claim, the Commission did not determine what is indispensable for the defence of the provider of X. Conversely, it was for the Specified External Advisers to put forward a reasoned request and provide for the reasons for which they considered the additional direct access to the addressee indispensable. They failed to demonstrate such indispensability in relation to the document for which they sought access as their request was overly broad. Nevertheless, in order to avoid a disproportionate delay or administrative burden the Commission granted access to 33 documents under Article 5(11) of Implementing Regulation (EU) 2023/1201, even though in its request under Article 5(6) of Implementing Regulation (EU) 2023/1201 the Specified External Advisers did not demonstrate why the documents to which access was sought were indispensable for the addressee's right to be heard. The late provision of one document that should have been disclosed under Article 5(2) of Implementing Regulation (EU) 2023/1201 was an isolated clerical error that the Commission promptly rectified.

(572) In the third place, as regards the provider of X's claim that the Commission's refusal to consider all their external advisers as Specified External Advisers was arbitrary,[606] it ignores the fact that the option to extend the list of initially Specified External Advisers is only available while the data room is open, as specified in Article 5(6) of Implementing Regulation (EU) 2023/1201. The Specified External Advisers were aware of these rules and failed to act within the required timeframe. It had been expressly asked beforehand to provide an exhaustive list of the specified external advisers, and the Terms of Disclosure Decision required that the specified external advisers be listed and sign the Data Room Rules, committing them to confidentiality obligations and procedural safeguards. Nevertheless, the provider of X requested that, not only those external advisers who had signed the Data Room Rules and the acknowledgement of the Terms of Disclosure Decision before entering the data room, but also all those external advisers belonging to the same law firm of the Specified External Advisers, should be able discuss the documents in the Commission's file, which were made accessible securely by the Commission in the data room.

606 Reply to the Preliminary Findings, paragraph 77.

(573) In the fourth place, the provider of X's suggestion that the Commission is '*learning by doing*' and failing to apply a lenient approach to the provider of X's compliance obligations is unfounded. The procedural timeline was clearly communicated, and the provider of X was given multiple opportunities to engage with the Commission regarding its obligations. The Commission's actions were guided by legal and procedural requirements, rather than by any intent to penalise the provider of X unfairly. The procedural framework was applied consistently, ensuring that the provider of X had a fair opportunity to exercise their rights of defence.

(574) In conclusion, the Commission's approach to access to file, the data room procedure, and the Article 5(6) Implementing Regulation Request was in compliance with Regulation (EU) 2022/2065 and Implementing Regulation (EU) 2023/1201. The provider of X's claims of unfairness, procedural breaches, and arbitrary decisions are unsubstantiated by the facts and the legal framework governing the proceedings.

8. DURATION OF THE PROVIDER OF X'S NON-COMPLIANCE

8.1. Non-compliance with Article 25(1) of Regulation (EU) 2022/2065

(575) The Commission finds that the provider of X's non-compliance with Article 25(1) of Regulation (EU) 2022/2065 which the Commission has established in sections 4.3 and 4.5 above commenced on the date when that provision became applicable to it that is on 28 August 2023 which is four months after the notification of the Designation Decision.

(576) On 5 June 2025, the Commission observed that the provider of X started to display a one-off disclaimer to recipients of its service at the top of the 'For You' timeline when loading the main website, reiterating the same information about the meaning of the blue-ribbon checkmark it displays on its help pages to which the disclaimer links as well.[607] On 30 September 2025, the provider of X informed the Commission that *'[a]s the Commission is aware, from 18 May 2025 to 23 September 2025, X displayed a prominent banner to all existing and new users in the EEA to ensure, once again, maximum possible user awareness about our blue check and verification policies.*' According to the provider of X *'[t]he banner was translated into all EEA languages and presented to users accordingly. During the period it was displayed, the banner received over 90 million unique impressions.*'

(577) While the additional one-off disclaimer on top of the "For You" timeline provides additional information to the recipients of X, the Commission considers that it does not appear to alter the misleading online interface design recipients continue to encounter in relation to the use of the 'verified' status feature, in particular because the disclaimer in question was displayed by the provider of X only for a limited period of time and it was shown to recipients only once when they logged in to the service. Therefore, the Commission considers that the aforementioned one-off disclaimer does not *prima facie* appear to terminate the infringement of Article 25(1) of Regulation (EU) 2022/2065, as set out in Section 4 of this Decision.

(578) The Commission therefore finds that the provider of X's non-compliance with Article 25(1) of Regulation (EU) 2022/2065 is still ongoing at the date of adoption of this Decision.

---

607 X website, https://x.com, accessed on 5 June 2025 (DSA.100101, Doc ID 278).

8.2. Non-compliance with Article 39 of Regulation (EU) 2022/2065

(579) The Commission finds that the provider of X's non-compliance with Article 39(1) of Regulation (EU) 2022/2065 which the Commission has established in section 5.3 and 5.5 above commenced on the date when that provision became applicable to it, that is on 28 August 2023 which is four months after the notification of the Designation Decision. By the date of adoption of this Decision, the provider of X has not introduced any material changes to the measures described in section 5.2 above. The Commission therefore concludes that the aforementioned non-compliance is still ongoing at the date of adoption of this Decision.

8.3. Non-compliance with Article 40(12) of Regulation (EU) 2022/2065

(580) The Commission finds that the provider of X's non-compliance with Article 40(12) of Regulation (EU) 2022/2065 which the Commission has established in section 5.3 and 5.5 above commenced on the date when that provision became applicable to X that is 28 August 2023 which is four months after the notification of the Designation Decision. By the date of adoption of this Decision, the provider of X has not introduced any material changes to the measures described in section 6.2 above. The Commission therefore concludes that the aforementioned non-compliance is still ongoing at the date of adoption of this Decision.

9. MEASURES TO ENSURE COMPLIANCE WITH THIS DECISION

9.1. Principles

(581) Pursuant to Article 73(3) of Regulation (EU) 2022/2065, where the Commission finds that the provider of the very large online platform or of the very large online search engine concerned does not comply with one or more of the relevant provisions of Regulation (EU) 2022/2065, it shall by decision order that provider to take the necessary measures to ensure compliance.

(582) A decision pursuant to Article 73(1) of Regulation (EU) 2022/2065 shall include an order to '*take the necessary measures to ensure compliance with the decision pursuant to* [Article 73] *paragraph 1 within a reasonable period specified therein and to provide information on the measures that the provider intends to take to comply with the decision*'.

(583) Pursuant to Article 73(4) of Regulation (EU) 2022/2065, the provider of the very large online platform or of the very large online search engine concerned shall provide the Commission with a description of the measures it has taken to ensure compliance with the non-compliance decision pursuant to paragraph (1) of that Article upon their implementation.

(584) Pursuant to Article 75(1) of Regulation (EU) 2022/2065, when adopting a decision pursuant to Article 73 of that Regulation in relation to an infringement by a provider of a very large online platform or of a very large online search engine of any of the provisions of section 5 of Chapter III, the Commission shall make use of the enhanced supervision system laid down in that Article.

(585) Pursuant to Article 75(2) of Regulation (EU) 2022/2065, the Commission shall require the provider of the very large online platform or of the very large online search engine concerned to draw a detailed action plan to remedy any effect of the infringement for the future and communicate such action plan, within a reasonable period specified by the Commission, to the Digital Services Coordinators, the Commission and the Board.

(586) Pursuant to Article 75(3) of Regulation (EU) 2022/2065, the Commission, taking into account the opinion of the Board, shall decide whether the measures set out in the action plan are sufficient to terminate or remedy the infringement, taking also into account whether adherence to relevant code of conduct is included among the measures proposed. The Commission shall also subsequently monitor measures taken by the provider of a very large online platform or of a very large online search engine concerned in the implementation of the action plan, taking into account also an independent audit of the provider. Pursuant to Article 75(4) of Regulation (EU) 2022/2065 if, following the implementation of the action plan, the Commission still considers that the infringement has not been fully remedied, or if the action plan has not been provided or is not considered suitable, the Commission shall be able to use any investigative or enforcement powers pursuant to Regulation (EU) 2022/2065, including the power to impose periodic penalty payments and initiating the procedure to disable access to the infringing service.

9.2. Application to this case

9.2.1. Measures to comply with Article 25(1) of Regulation (EU) 2022/2065

(587) As set out in recital 577 above, the Commission considers that the changes implemented by the provider of X between 18 May and 23 September 2025 to the measure described in Section 4.2 do not prima facie appear to terminate that provider's non-compliance with Article 25(1) of Regulation (EU) 2022/2065. It is therefore necessary for the provider of X to bring the non-compliance with Article 25(1) of Regulation (EU) 2022/2065 established in section 4 of this Decision to an end within 90 days from the notification of this Decision. It is also necessary to ensure that the provider of X refrains from repeating any conduct having the same or an equivalent object or effect to the non-compliance established in this Decision.

(588) In order to ensure compliance with Article 25(1) of Regulation (EU) 2022/2065, the Commission considers, pursuant to Article 73(3) of that Regulation, that the provider of X should ensure that X's online interface does not deceive recipients of that service as to the identity, authenticity and notability of accounts carrying the blue-ribbon checkmark. To that end, the Commission considers that the following changes should be made to the measures described in Section 4.2:

(a) refraining from textually claiming that users are verified when no reliable verification process has in fact taken place;

(b) more prominently and directly providing clear and non-misleading information about the meaning of the blue-ribbon checkmark in X's online interface;

(c) refraining from misleading algorithmic amplification of accounts carrying the blue-ribbon checkmark; and

(d) publicly announcing and promoting to recipients of X's service these changes to the blue-ribbon checkmark to ensure transparency.

(589) Pursuant to Article 73(3) of Regulation (EU) 2022/2065 the provider of X shall provide the Commission, with a description of the measures it intends to take to ensure compliance with this Decision, and in particular to terminate or remedy tits non-compliance with Article 25(1) of Regulation (EU) 2022/2065 identified in section 4 of this Decision within 60 working days from the date of notification of this Decision and bring that infringement effectively to end within 90 working days from the date of notification of this Decision.

(590) In addition to terminating the infringement identified in section 4 of this Decision, the provider of X should implement measures that prevent the recurrence of that infringement. When submitting its proposed measures, the provider of X should ensure that its proposal is sufficiently reasoned and detailed to enable the Commission to assess whether the proposed measures will ensure that the recurrence of the non-compliance is avoided and thus that it will be effectively brought to an end.

9.2.2. Measures to comply with Article 39 of Regulation (EU) 2022/2065

(591) It is necessary for the provider of X to bring its non-compliance with Article 39 of Regulation (EU) 2022/2065 established in section 5 of this Decision to an end under the procedure set out in Article 75 of that Regulation. It shall also refrain from repeating any act of conduct having the same or equivalent object or effect to the non-compliance established in this Decision.

(592) In order to ensure compliance with Article 39 of Regulation (EU) 2022/2065, the Commission considers that the provider of X should ensure that X's advertisement repository includes a searchable and reliable tool that allows meaningful multicriteria queries, including through fully functional and effective APIs, allowing supervision and research into emerging risks by the distribution of advertising online. To that end, the Commission considers that the action plan that the provider of X will be obliged to draw up and communicate to the Commission, in accordance with Article 75(2) of Regulation (EU) 2022/2065 and as specified in recital 597 below, should consider the following changes to the measures described in Section 5.2:

(a) changing the design and implementation of X's advertisement repository in a way that allows multicriteria queries based on at least all the information related to advertisements listed in Article 39(2) of Regulation (EU) 2022/2065;

(b) providing search results in a way that they are accessible in a specific section of X's online interface without relying on a third-party software;

(c) changing the design and implementation of X's advertisement repository in a way that the response time of the online interface is reduced such that the tool is searchable;

(d) changing the design and implementation of X's advertisement repository in a way that the search results provided by the search tool are reliable and provide at least all the information listed in Article 39(2) of Regulation (EU) 2022/2065 for the entire period during which they present an advertisement and until one year after the advertisement was presented for the last time on the provider of X's online interface; and

(e) changing the design and implementation of X's advertisement repository in a way that there are no technical or contractual limitations that prevent the availability of information on that repository through APIs.

9.2.3. Measures to comply with Article 40(12) of Regulation (EU) 2022/2065

(593) It is necessary for the provider of X to bring its non-compliance with Article 40(12) of Regulation (EU) 2022/2065 established in section 6 of this Decision to an end under the procedure set out in Article 75 of that Regulation. It shall also refrain from repeating any act of conduct having the same or equivalent object or effect to the non-compliance identified in this Decision.

(594) In order to ensure compliance with Article 40(12) of Regulation (EU) 2022/2065, the Commission considers, pursuant to Article 73(3) of that Regulation, that the provider

of X should take all the necessary steps to give Qualified Researchers access without undue delay to data, including where technically possible, real-time data, that is publicly accessible in X's online interface in a manner that effectively enables Qualified Researchers to carry out research that contributes to the detection, identification and understanding of systemic risks in the Union pursuant to Article 34(1). To that end, the Commission considers that the action plan that the provider of X will be obliged to draw up and communicate to the Commission, in accordance with Article 75(2) of Regulation (EU) 2022/2065 and as specified in recital 597 below, should consider the following changes to the measures described in Section 6.2:

(a) refraining from imposing eligibility requirements preventing Qualified Researchers from accessing data that is publicly accessible in X's online interface, in particular by:

(1) refraining from rejecting data access applications without assessing them based on an adequate and up-to-date expertise of the state of the art of research contributing to the detection, identification and understanding of systemic risks in the Union pursuant to Article 34(1) of Regulation (EU) 2022/2065;

(2) refraining from imposing eligibility requirements related to the institutional affiliation and geographic location of researchers which go beyond the requirements of Article 40(12) of Regulation (EU) 2022/2065;

(b) giving Qualified Researchers free-of-charge access to the X API without undue delay, including by refraining from aligning the timelines of applications submitted on different dates;

(c) giving Qualified Researchers access without undue delay to the volumes of data that are necessary and proportionate for the purposes of their research, including by stating publicly the standard API quotas granted to successful applicants by default, thus enabling applicants to request extensions of these quotas in their initial applications, and by swiftly assessing the need for such extensions when they are requested;

(d) refraining from contractually prohibiting Qualified Researchers from independently accessing data that is publicly accessible in X's online interface, such as by means of data scraping, including by adapting the wording of X's terms of service.

9.2.4. Enhanced supervision system of remedies to address infringements of obligations laid down in section 5 of Chapter III of Regulation (EU) 2022/2065

(595) In accordance with Article 73(4) of Regulation (EU) 2022/2065, the provider of X shall provide the Commission with a description of the measures it has taken to ensure compliance with Articles 25, 39 and 40(12), as set out in sections 4, 5 and 6 of this Decision, upon their implementation.

(596) In addition, in accordance with Article 75(2) of Regulation (EU) 2022/2065, the provider of X shall draw up and communicate to the Digital Services Coordinators, the Commission, and the Board, an action plan setting out the necessary measures which are sufficient to terminate or remedy the infringements of Articles 39 and 40(12) of that Regulation. Those measures shall include a commitment to perform an independent audit in accordance with Article 37(3) and (4) of Regulation (EU)

2022/2065 on the implementation of the other measures, and shall specify the identity of the auditors, as well as the methodology, timing and follow-up of the audit.

(597) The aforementioned action plan shall be communicated to the Digital Services Coordinators, the Commission, and the Board within 90 working days as from receipt of this Decision.

10. FINES

10.1. Principles

(598) Pursuant to Article 74(1), point (a), of Regulation (EU) 2022/2065, the Commission may impose by decision fines on the provider of a very large online platform or of a very large search engine where it finds that that provider intentionally or negligently fails to comply with the relevant provisions of that Regulation. Fines under Regulation (EU) 2022/2065 are ultimately aimed at ensuring effective compliance by providers of very large online platforms or of very large search engines with their obligations under applicable laws and the prevention of non-compliant conduct that could undermine the objectives of applicable law (i.e., in the present case Regulation (EU) 2022/2065).

(599) Pursuant to Article 74(4) of Regulation (EU) 2022/2065, in fixing the amount of the fine, the Commission shall have regard to the nature, gravity, duration, and recurrence of the infringement. In general, fines shall be effective, proportionate, and dissuasive.[608] The Commission must ensure that any aggravating or mitigating circumstances are also reflected in the fines imposed. In doing so, the Commission must set the fines at a level sufficient to ensure deterrence.

(600) Any fine imposed by the Commission shall not exceed 6% of the provider's total worldwide annual turnover in the preceding financial year. Pursuant to Article 3, point x, of Regulation (EU) 2022/2065, the notion of turnover is to be understood as the amount derived by an undertaking within the meaning of Article 5(1) of Council Regulation (EC) No 139/2004. Additionally, the Commission must ensure that its fines are compatible with the principles of equal treatment and proportionality,[609] while having, at the same time, the necessary deterrent effect.[610]

10.2. Intent or negligence

(601) As mentioned in recital 598 above, a fine may be imposed on a provider of a very large online platform or of a very large search engine that '*intentionally or negligently*' infringes the relevant provisions of Regulation (EU) 2022/2065.

10.2.1. The Commission's Preliminary Findings

(602) In its Preliminary Findings, the Commission found that the conduct described in sections 4.2, 5.2 and 6.2 above constitutes three separate infringements of Regulation (EU) 2022/2065 and that, therefore, fines should be imposed on the provider of X for

608 Judgment of 8 July 2020, *VQ v European Central Bank*, Case T 203/18, EU:T:2020:313, Judgment of 8 July 2020, *CA Consumer Finance v European Central Bank*, Case T-578/18, EU:T:2020:306 and Judgment of 8 July 2020, *Crédit Agricole v European Central Bank*, Case T-576/18, EU:T:2020:304.

609 See, for instance and by analogy, Judgment of 9 September 2015, Case T-92/13, *Philips v Commission*, EU:T:2015:605, paragraph 194 and the case-law cited therein.

610 See, for instance and by analogy, Judgment of 13 July 2011, Case T-59/07, *Polimeri Europa v Commission*, EU:T:2011:361, paragraph 243 and the case-law cited therein.

each of these infringements separately in accordance with Article 74(1), point (a), of Regulation (EU) 2022/2065.

(603) As follows from the descriptions in sections 4.3, 5.3 and 6.3 above, the provider of X has acted intentionally or at the very least negligently in relation to each of the infringements identified in those sections. The provider of X could not have been unaware that the measures described in sections 4.2, 5.2 and 6.2 above do not, at first sight, comply with the requirements laid down in Articles 25(1), 39 and 40(12) of Regulation (EU) 2022/2065 respectively. Despite clear obligations stemming from these provisions and the fact that the provider of X is comprised of large entities with experience in compliance and capable of asking for legal advice, the provider of X took the conscious decisions to adopt the measures described in sections 4.2, 5.2 and 6.2 above, which fall short of its legal obligations under Regulation (EU) 2022/2065.

10.2.2. The arguments of the provider of X

(604) In its Reply to the Preliminary Findings on the matter of intent or negligence, the provider of X does not specifically address each non-compliance preliminarily established in the Preliminary Findings, but rather argues in general that Articles 25(1), 39 and 40(12) of Regulation (EU) 2022/2065 do not contain clear obligations, at least in respect of the alleged infringements covered by the Preliminary Findings. In particular, the provider of X argues that: (i) the wording of these provisions is often vague; (ii) the Commission stretches the provisions *'well beyond their natural meaning*'; (iii) there is an absence of case-law or decisional practice; and (iv) the Commission has provided no substantive guidance.[611] The provider of X contests that it has relevant compliance experience that would justify a fine as Regulation (EU) 2022/2065 is a new tool.[612] The provisions at stake would allow for different interpretations and compliance solutions, as confirmed by the Audit Report for X.[613] The provider of X further argues that it has '*always acted in good faith*' and continues to have '*all the reasons to believe*' that its practices were in full compliance with Regulation (EU) 2022/2065.[614] Even if the provider of X had acted intentionally or negligently (quod non), the novelty of the instrument should refrain the Commission from imposing fines.[615]

10.2.3. The Commission's assessment of the provider of X's arguments

(605) The arguments put forward by the provider of X do not call into question the Commission's preliminary finding that the provider of X acted intentionally or at the very least negligently in relation to the infringements identified in this Decision.

(606) The provider of X's unfounded claims that the obligations at stake lack clarity and that the Commission has failed to provide guidance have been addressed in sections 4.5, 5.5 and 6.5 above. As regards the novelty of Regulation (EU) 2022/2065, it in no way excludes the intentional or at the very least negligent behaviour of the provider of X. In any event, compliance with new legislation requires regulated entities to carefully assess their new obligations and to ensure compliance with them from the moment they start to apply. If the provider of X's argument on this point were accepted, it

611 Reply to the Preliminary Findings, paragraph 348 (DSA.100101, Doc ID 233; DSA.100102, Doc ID 337-2; DSA.100103, Doc ID 289-2).

612 Reply to the Preliminary Findings, paragraph 349.

613 Reply to the Preliminary Findings, paragraph 350.

614 Reply to the Preliminary Findings, paragraph 352.

615 Reply to the Preliminary Findings, paragraph 356.

would effectively delay the full entry into application of Regulation (EU) 2022/2065, contrary to the will of the Union legislature to apply that regulation as quickly as possible to very large online platforms and very large online search engines.[616] While the novelty of Regulation (EU) 2022/2065 can be taken into account as a mitigating factor when setting the level of fines, this circumstance alone does not bar the Commission from imposing fines in the first place. The Audit Report for X also does not change the Commission's conclusions on the provider of X's intent or at the very least negligence. As pointed out in recital 246 above, the purpose of the annual audit is to contribute and inform the Commission in its supervision and enforcement activities under Regulation (EU) 2022/2065. However, this does not mean that the Commission must accept at face-value the outcome of such an audit. particularly when an audit does not include conclusions on obligations that are covered by the Commission's proceedings. Lastly, the provider of X has close to two decades of experience in providing researchers access to publicly accessible data, so it should not have been difficult for it to assess what measures would ensure compliance with Article 40(12) of Regulation (EU) 2022/2065.[617]

(607) It follows that the provider of X could not have been unaware that it was infringing Articles 25(1), 39 and 40(12) of Regulation (EU) 2022/2065 and that it acted intentionally or, at the very least, negligently in doing so.

10.3. Calculation of the fine

10.3.1. Nature of the infringements

10.3.1.1. The Commission's Preliminary Findings

(608) In its Preliminary Findings, the Commission found that the nature of the infringements of Articles 25(1), 39 and 40(12) of Regulation (EU) 2022/2065 by the provider of X is particularly serious.

(609) As regards Article 25(1) of Regulation (EU) 2022/2065, while the provider of X significantly weakened the verification process that the provider of Twitter had initially put in place for users to obtain a 'verified' status, it intentionally maintained the design features of the 'verified' status put in place by the provider of Twitter so as to deceive the recipients of X in relation to the identity, notability and authenticity of 'verified' accounts.[618] As regards Article 39 of Regulation (EU) 2022/2065, the provider of X intentionally designed and implemented the search tool of its advertisement repository in such a way that significantly weakens the searchability of that repository and its practical usefulness.[619] As regards Article 40(12) of Regulation (EU) 2022/2065, the provider of X took the deliberate decision to dismantle the Twitter Academic API program, which was available to academic researchers until the spring of 2023, without providing for an adequate replacement, and failed to give Qualified Researchers access to data that is publicly accessible in X's online interface.

10.3.1.2. The arguments of the provider of X

(610) In its Reply to the Preliminary Findings, the provider of X claims that the Commission's preliminary finding of 'particular seriousness' regarding the nature of

616 See Order in Case C-639/23 P(R), *Commission v Amazon Services Europe*, EU:C:2024:277, paragraph 162.

617 See in particular recitals 291, 422 to 427, 433 above.

618 See recitals 97 to 102 above.

619 See recitals 177 to 212 above.

the infringements lacks context and quantification.[620] The provider of X argues that the Commission merely described the infringements which would be insufficient to assess their nature.[621] The provider of X also states that there is no internal ranking of seriousness within Regulation (EU) 2022/2065 and that the Preliminary Findings do not argue that the allegedly infringed substantive provisions of the DSA are somehow more important than others.[622]

10.3.1.3. The Commission's assessment of the provider of X's arguments

(611) Contrary to what the provider of X claims, the Commission's assessment of the nature of the infringements does not lack context or quantification. Article 74 of Regulation (EU) 2022/2065 is inspired by Article 23 of Regulation 1/2003 and the decisional practice of the Commission in antitrust cases. Consequently, when determining the appropriate amount of fine to be imposed on providers of very large online platforms and of very large online search engines for infringements of Regulation (EU) 2022/2065, the Commission relies by analogy on a well-established methodology construed by Union courts in the field of Union competition law, including fines for infringements of the obligations set out in Regulation 139/2004 on the control of concentrations between undertakings.[623] Those courts have already clarified that, in the absence of guidelines, the Commission fulfils its obligation to state reasons by showing clearly and unequivocally the factors taken into account when imposing a fine.[624] The Commission recalls in this respect that the seriousness of an infringement has been considered as a parameter to ascertain the nature of an infringement in its decisional practice in antitrust proceedings.[625]

(612) The assessment of seriousness is a qualitative, rather than quantitative, assessment and it plays in relation to establishing both the nature and the gravity of the infringements. In this context, by describing the specific infringements the Commission illustrated the manner in which they negatively affect the objectives pursued by Regulation (EU) 2022/2065. Regulation (EU) 2022/2065 is a central element of the policy developed by the Union legislature in the digital sector. In the context of that policy, Regulation (EU) 2022/2065 pursues objectives of great importance, since it seeks, as is apparent from Article 1 thereof, to contribute to the proper functioning of the internal market and to ensure a safe, predictable and trusted online environment that facilitates innovation and in which fundamental rights enshrined in the Charter are effectively protected.[626] With Regulation (EU) 2022/2065, the Union legislature considered that very large online platforms and very large online search engines play an important role

620 Reply to the Preliminary Findings, paragraph 357.

621 Reply to the Preliminary Findings, paragraph 359.

622 Reply to the Preliminary Findings, paragraph 359.

623 Council Regulation (EC) No 139/2004 of 20 January 2004 on the control of concentrations between undertakings ("Merger Regulation") OJ L 24, 29.1.2004, pp. 1–22.

624 *See* judgements of 18 May 2022, Canon v. Commission, T-609/19, EU:T:2022:299, paragraph 435 and the case-law cited; and of 17 December 2014, Pilkington Group and Others v Commission (T-72/09, not published, EU:T:2014:1094, paragraphs 247 and 248 and the case-law cited).

625 *See e.g.*, Case AT.40178 – CAR EMISSIONS, paragraph 223; Case M.8181 MERCK / SIGMA-ALDRICH, paragraph 473; Case M.7993 - ALTICE / PT PORTUGAL, paragraph 573; Case M.8228 - FACEBOOK / WHATSAPP, paragraph 97; Case M.4994 - ELECTRABEL / COMPAGNIE NATIONALE DU RHONE, paragraph 188.

626 Order of the Vice President of the Court of 27 March 2024 in Case C-639/23 P(R), European Commission vs. Amazon Services Europe, EU:C:2024:277, paragraph 155. See also Order of the President of the General Court of 2 July 2024 in Case T-138/24 R, Aylo Freesites Ltd. vs. European Commission, EU:T:2024:431, paragraph 113.

in the digital environment and that they may give rise to risks for society which differ, in terms of their scale and impact, from those attributable to smaller online platforms.[627]

(613) Deceiving recipients of X in relation to the identity, notability and authenticity of 'verified' accounts undermines the effectiveness of Regulation (EU) 2022/2065. Article 25(1) of Regulation (EU) 2022/2065 constitutes one of the few express prohibitions of specific behaviour in that Regulation. This prohibition is justified by the harmfulness of dark patterns for the autonomy, decision-making, or choice of the recipients of the service, which is detrimental to a safe, predictable and trusted online environment. While Article 25(1) of Regulation (EU) 2022/2065 is addressed to all providers of online platforms, its infringement by the provider of a very large online platform has direct negative effects on a particularly large number of recipients of the service. The resulting uninformed decisions of those recipients can further compound the harm. This direct harm to recipients of the service increases the seriousness of the nature of the infringement of Article 25(1) of Regulation (EU) 2022/2065.

(614) Designing and implementing the search tool of X's advertisement repository in such a way that significantly weakens the searchability of that repository and its practical usefulness equally undermines the effectiveness of Regulation (EU) 2022/2065. An infringement of Article 39 of Regulation (EU) 2022/2065 negatively affects the transparency on the advertisement practices of very large online platforms and of very large online search engines. As clarified in Recital 95 of Regulation (EU) 2022/2065, advertising systems used by very large online platforms and very large online search engines pose particular risks and require further public and regulatory supervision on account of their scale and ability to target and reach recipients of the service based on their behaviour within and outside that platform's or search engine's online interface.[628] The resulting lack of transparency impairs public scrutiny of the advertising market and gives the provider of X an undue advantage as compared to those providers of very large online platforms and of very large online search engines that comply with Article 39 of Regulation (EU) 2022/2065. While Article 39 of Regulation (EU) 2022/2065 is of a procedural nature and its infringement causes indirect harm to recipients of the service, it causes direct harm to those involved in research and supervision.

(615) Rejecting the applications of Qualified Researchers, failing to give Qualified Researchers adequate access without undue delay, and contractually prohibiting Qualified Researchers from independently accessing publicly accessible data also undermine the effectiveness of Regulation (EU) 2022/2065. An infringement of Article 40(12) of Regulation (EU) 2022/2065 hinders research that contributes to the detection, identification, and understanding of systemic risks in the Union. The potential harm resulting from the stifling of legitimate research projects can negatively impact the ability to study and ultimately address such systemic risks. The increased risk profile of very large online platforms and of very large online search engines warrants that they are subject to this increased scrutiny by Qualified Researchers.

627 Order of the Vice President of the Court of 27 March 2024 in Case C-639/23 P(R), European Commission vs. Amazon Services Europe, EU:C:2024:277, paragraph 159.

628 Order of the Vice President of the Court of 27 March 2024 in Case C-639/23 P(R), European Commission vs. Amazon Services Europe, EU:C:2024:277, paragraph 160. See also Order of the President of the General Court of 2 July 2024 in Case T-138/24 R, Aylo Freesites Ltd. vs. European Commission, EU:T:2024:431, paragraph 117.

Recital 96 of Regulation (EU) 2022/2065 clarifies that the Union legislature considered that "*[i]nvestigations by researchers on the evolution and severity of online systemic risks are particularly important for bridging information asymmetries and establishing a resilient system of risk mitigation*", informing providers, regulators and the public. The objective of this provision is to offer transparency, an overall goal of Regulation (EU) 2022/2065, and scrutiny over the risk management obligations of providers of very large online platforms and of very large online search engines. An infringement of Article 40(12) of Regulation (EU) 2022/2065 undermines the effectiveness of that Regulation as it hampers effective research into the systemic risks in the Union pursuant to Article 34(1) of that Regulation, which can in turn inform and advance the public debate around the risks that stem from the services of very large online platforms and of very large online search engines and the supervisory work of the regulators. An infringement of Article 40(12) of Regulation (EU) 2022/2065 by rejecting the applications of Qualified Researchers and, at the very least, fostering ambiguity regarding the application of a contractual prohibition of independent data access to Qualified Researchers, while knowingly maintaining an application review process that regularly lead to undue delays, frustrates the Union legislature's objective to effectively enable researchers to contribute to a resilient system of risk management. While Article 40(12) of Regulation (EU) 2022/2065 is of a procedural nature and its infringement therefore causes indirect harm to recipients of the service, it causes direct harm to those involved in research and supervision.

(616) Based on the aforementioned considerations, the Commission finds that the nature of the infringement of Article 25(1) of Regulation (EU) 2022/2065 by the provider of X is particularly serious as it violates the prohibition of a specific behaviour that causes direct harm to the recipients of the service and undermines the effectiveness of that Regulation, while the nature of the infringements of Articles 39 and 40(12) of Regulation (EU) 2022/2065 by the provider of X are at least serious as they undermine the effectiveness of that Regulation.

10.3.2. Gravity of the infringements

10.3.2.1. The Commission's Preliminary Findings

(617) In its Preliminary Findings, the Commission found that the gravity of the infringements of Articles 25(1), 39 and 40(12) of Regulation (EU) 2022/2065 by the provider of X is particularly serious and that such infringements can undermine the objectives and effectiveness of that Regulation.

(618) As regards Article 25(1) of Regulation (EU) 2022/2065, the provider of X deceptively designed, organised and operated the 'verified' status for X, an online platform with particular relevance for the political discourse and at a time with multiple consequential elections in the Union. As regards Article 39 of Regulation (EU) 2022/2065, the provider of X intentionally chose to design its advertisement repository in a manner that significantly weakens its searchability and its practical usefulness, thus frustrating the legislature's pursuit of its objective. As regards Article 40(12) of Regulation (EU) 2022/2065, the provider of X has prevented oversight and scrutiny of the systemic risks posed by X in the Union by rejecting the applications of eligible researchers, failing to give Qualified Researchers adequate access without undue delay, and contractually prohibiting Qualified Researchers from independently accessing publicly accessible data.

(619) For all three infringements, their gravity is increased by the widespread use of X in the Union and the fact that the relevant conduct by the provider of X was the same across

the entire Union. The number of active recipients of X in the Union is more than the double the threshold for designation as a very large online platform or as a very large search engine within the meaning of Article 33(1) of Regulation (EU) 2022/2065 (*i.e.*, 45 million active recipients). In fact, X is a social media platform with one of the widest reaches in the Union, as it had over 109 million (109 191 304) average monthly active recipients in the Union and, according to X's transparency report of April 2024, a significant user base proportionate to the population of each Member State in the majority of Member States at the time the Preliminary Findings were adopted.[629] X not only functions as an important source of news and entertainment for its users who upload content and interact with each other, it also offers a relevant forum for political dialogue. Therefore, the infringements of the aforementioned provisions of Regulation (EU) 2022/2065 by the provider of X have a serious impact on a large number of recipients of the service in the Union.

10.3.2.2. The arguments of the provider of X

(620) In its Reply to the Preliminary Findings, the provider of X argues that the Commission simply listed the general objectives of Regulation (EU) 2022/2065 and did not explain how the infringements undermine its effectiveness.[630] The provider of X also considers that the number of users in the Union is not an appropriate reference to assess the gravity of the infringements because this reasoning would presumably apply to '*any provider*' covered by the substantive obligations of Regulation (EU) 2022/2065.[631]

10.3.2.3. The Commission's assessment of the provider of X's arguments

(621) Contrary to what the provider of X claims, the impact of the infringements of Articles 25(1), 39 and 40(12) of Regulation (EU) 2022/2065 on the effectiveness of that Regulation are of central importance for setting the level of fines. The Commission has already addressed these aspects above in relation to the nature of the infringements. The gravity of the infringement is closely linked to the nature of the infringement and should be read in combination with the considerations set out in the preceding paragraphs. For the reasons explained in section 10.3.1.3 of this Decision the Commission finds that the nature of the infringement of Article 25(1) of Regulation (EU) 2022/2065 by the provider of X is particularly serious while the nature of the infringements of Articles 39 and 40(12) of Regulation (EU) 2022/2065 by the provider of X are at least serious as they undermine the effectiveness of that Regulation.

(622) In addition, the infringements of Articles 25(1), 39 and 40(12) of Regulation (EU) 2022/2065 by the provider of X are multifaceted and intentional or, at the very least, negligent. The infringement of Article 25(1) of Regulation (EU) 2022/2065 identified in section 4 of this Decision deceives recipients of the X service in a manner that materially distorts or impairs their ability to make free and informed decisions when they are active online, and not only on the X online platform, thus impacting a central objective of that regulation, namely ensuring a safe, predictable and trusted online environment. The infringement of Article 39 of Regulation (EU) 2022/2065 identified in section 5 of this Decision leads to a lack of advertisement transparency in relation to

629 X, DSA Transparency Report – April 2024, https://transparency.x.com/dsa-transparency-report-2024-april.html, accessed on 7 May 2025 (DSA.100101, Doc ID 195; DSA.100102, Doc ID 258; DSA.100103, Doc ID 222).

630 Reply to the Preliminary Findings, paragraph 360.

631 Reply to the Preliminary Findings, paragraph 360.

X. The infringement of Article 40(12) of Regulation (EU) 2022/2065 identified in section 6 of this Decision harms the oversight and scrutiny of a wide range of systemic risks in the Union in relation to X. The resulting negative impact on the supervision and enforcement of the provider of X's compliance with the obligations in Articles 34 and 35 of Regulation (EU) 2022/2065 increases the seriousness of the infringement of Article 40(12) of that Regulation.

(623) The reach of a service is a relevant element when assessing the gravity of the infringements at stake, as it is an indicator of the level of due diligence that a provider should exercise for its service and of the magnitude of the harm caused by an infringement of that provider's obligations under Regulation (EU) 2022/2065. As already stated in the Preliminary Findings, X is one of the largest social media platforms in the Union with a significant user base across the Union. The service constitutes an important source of news and entertainment and a relevant forum for political dialogue. Even among the very large online platforms designated by the Commission, X has a particularly large user base within the Union that by far surpasses the threshold of 45 million average monthly active recipients. X's transparency reports of October 2024 and April 2025 indicate similar numbers, amounting respectively to over 105 million (105 271 027) and 94 million (94 830 300) average monthly active recipients in the Union.[632] A higher reach also increases the relevance of advertisement transparency and independent research on the risks posed by a specific service. Additionally, while Article 25(1) of Regulation (EU) 2022/2065 applies to all providers of online platforms, irrespective of their number of recipients, the gravity of the infringement increases where more users are deceived and thereby harmed through the actions of a provider. This harm to a particularly large user base within the Union and the potential amplification of posts by malicious actors to a large audience increases the seriousness of the gravity of the infringement of Article 25(1) of Regulation (EU) 2022/2065.

(624) When assessing the gravity of the infringement, the Commission also needs to consider the deterrent effect of fines. The overall fines must have a sufficient deterrent effect in relation to the provider of X, but also in relation to other regulated services to avoid future breaches of Regulation (EU) 2022/2065. In the case of an undertaking of the size and financial capacity of the single economic entity ultimately controlled by Mr. Elon Musk,[633] the amount of the fine must be significant to ensure a deterrent effect.[634] Additionally, the potential economic benefits that the provider of X could draw from non-compliance with Articles 25(1), 39 and 40(12) of Regulation (EU)

---

632 X, DSA Transparency Report – October 2024 and DSA Transparency Report – April 2025, https://transparency.x.com/en/reports/dsa-transparency-report, accessed on 7 May 2025 (DSA.100101, Doc ID 295; DSA.100102, Doc ID 388; DSA.100103, Doc ID 331).

633 *See* judgment of 13 February 2025, ILVA A/S, C-383/23, ECLI:EU:C:2025:84, paragraph 36: *'The concept of 'undertaking' must also be taken into account in order to assess the actual or material economic capacity of the recipient of the fine and thus to ascertain whether the fine is at the same time effective, proportionate and dissuasive.'*.

634 The EU Courts have established that the financial strength and size of the company may be taken into account when imposing a fine. *See e.g.*, judgement of 12 December 2012, Electrabel, T-332/09, paragraph 282: '*in determining the amounts of fines, the Commission is entitled to take into account the need to ensure that fines have a sufficient deterrent effect […]. Furthermore, the link between, on the one hand, undertakings' size and global resources and, on the other, the need to ensure that a fine has a deterrent effect cannot be denied. Accordingly, when the Commission calculates the amount of the fine it may take into consideration, inter alia, the size and the economic power of the undertaking concerned.*'.

2022/2065 must be duly considered. Such economic benefits can be seen in any revenues generated by making the Verified Account status available to paying subscribers, the potential undue increase of advertisement business due to a lack of advertisement transparency and due to the obstruction of Qualified Researchers' oversight and scrutiny of the systemic risks posed by X in the Union, as well as additional revenues generated through the fees charged to Qualified Researchers in the absence of free access to publicly accessible data under Article 40(12) of Regulation (EU) 2022/2065.

(625) Based on the above, in relation to the infringements of Articles 25(1), 39 and 40(12) of Regulation (EU) 2022/2065 by the provider of X identified in this Decision the Commission finds each of those infringements to be at least serious.

10.3.3. Duration of the non-compliance

10.3.3.1. The Commission's Preliminary Findings

(626) In its Preliminary Findings, the Commission found that the provider of X's infringements of Articles 25(1), 39, and 40(12) of Regulation (EU) 2022/2065 started on 28 August 2023, that is the date on which all the obligations stemming from the aforementioned provisions of Regulation (EU) 2022/2065 started to apply to that provider in relation to the provision of X in the Union. With regard to the infringement of Article 40(12) of Regulation (EU) 2022/2065, the fact that the additional tools described in recitals 294 to 298 above were put in place after 9 November 2023 does not alter the Commission's finding that the provider of X's infringement of Article 40(12) of Regulation (EU) 2022/2065 is ongoing. At the time of the adoption of the Preliminary Findings, the Commission considered all of the infringements to be ongoing.

10.3.3.2. The arguments of the provider of X

(627) The provider of X did not comment on the duration of the infringements beyond its general denial that such infringements occurred.

10.3.3.3. The Commission's assessment of the provider of X's arguments

(628) In light of the findings in section 8 above and given the lack of arguments put forward by the provider of X contesting the duration of the infringements, the Commission concludes that the provider of X's non-compliance with Articles 25(1), 39 and 40(12) of Regulation (EU) 2022/2065 lasted for at least 27 months.

10.3.4. Recurrence of the infringements

(629) The recurrence factor is not relevant in the present case, since this is the first decision addressed to the provider of X pursuant to Article 73(1) of Regulation (EU) 2022/2065. To date, no other decision pursuant to Articles 73(1) or 74 of Regulation (EU) 2022/2065 has been adopted in relation to the provider of X.

10.3.5. Aggravating circumstances

(630) The Commission considers that there are no aggravating circumstances in this case.

10.3.6. Mitigating circumstances

(631) The Commission considers that the novelty of Regulation (EU) 2022/2065 and the fact that the present Decision is the first non-compliance decision adopted under that Regulation constitutes a mitigating factor for calculating the fines imposed on the provider of X for the three infringements at stake. Regulation (EU) 2022/2065 has

established a new framework governing the provision of intermediary services in the internal market, including the obligations applicable to very large online platforms and to very large online search engines. The Commission considers that this should be taken into account when assessing the level of the fine for the provider of X's non-compliance with Articles 25(1), 39 and 40(12) of Regulation (EU) 2022/2065 at this point in time. The Commission therefore takes this factor into account when calculating the fine.

10.3.7. Conclusion

(632) In view of the above, the Commission considers that the provider of X has intentionally or, at the very least, negligently infringed Articles 25(1), 39 and 40(12) of Regulation (EU) 2022/2065. These infringements are of a serious nature in relation to Articles 39 and 40(12) of Regulation (EU) 2022/2065 and of a particularly serious nature in relation to Article 25(1) of that Regulation, and of a serious gravity and of a medium duration for all infringements. Finally, the Commission has taken into account the mitigating circumstance that this Decision is the first non-compliance decision adopted under Regulation (EU) 2022/2065.

10.4. The relevant turnover

(633) Pursuant to Article 74(1) of Regulation (EU) 2022/2065, where the Commission finds that a provider of a very large online platform or of a very large online search engine, intentionally or negligently, infringes the relevant provisions of this Regulation, the fine that it imposes for the infringement shall not exceed 6% of the total worldwide annual turnover of the provider in the preceding financial year. Article 3, point x, of Regulation (EU) 2022/2065 defines the notion of 'turnover' as the amount derived by an undertaking within the meaning of Article 5(1) of Council Regulation (EC) No 139/2004. Accordingly, Article 5(1) of Council Regulation (EC) No 139/2004, turnover shall *'comprise the amounts derived by the undertakings concerned in the preceding financial year from the sale of products and the provision of services falling within the undertakings' ordinary activities after deduction of sales rebates and of value added tax and other taxes directly related to turnover'*.

(634) Article 5(4) of Council Regulation (EC) No 139/2004 further elaborates on the notion of aggregate turnover of the undertaking concerned indicating that it '*shall be calculated by adding together the respective turnovers of the following: (a) the undertaking concerned; (b) those undertakings in which the undertaking concerned, directly or indirectly: (i) owns more than half the capital or business assets, or (ii) has the power to exercise more than half the voting rights, or (iii) has the power to appoint more than half the members of the supervisory board, the administrative board or bodies legally representing the undertakings, or (iv) has the right to manage the undertakings' affairs; (c) those undertakings which have in the undertaking concerned the rights or powers listed in (b); (d) those undertakings in which an undertaking as referred to in (c) has the rights or powers listed in (b); (e) those undertakings in which two or more undertakings as referred to in (a) to (d) jointly have the rights or powers listed in (b).*'

(635) By referring to the notion of turnover of the undertaking concerned employed in Council Regulation (EC) No 139/2004, Regulation (EU) 2022/2065 incorporates, for the purpose of imposing a fine, a concept already established by competition law as interpreted by the Union courts. This incorporation requires the Commission, when determining the total worldwide annual turnover of the provider for the 6% ceiling that apply when levying a fine, to consider that provider as an undertaking, including as a

single economic unit,[635] within the meaning of Articles 101 and 102 TFEU, as construed by the case law of the Union courts.[636]

(636) The provider of X argues in its Reply to the Preliminary Findings that only TIUC is the addressee of the Designation Decision and that therefore the only entity relevant for the purpose of calculating the ceiling of a potential fine is that subsidiary, just as it was the relevant entity for the determination of the supervisory fee.[637] That argument contradicts the explicit reference to the notion of 'turnover' in Article 3(x) of Regulation (EU) 2022/2065 and its reference to the notion of 'undertaking concerned' within the meaning of Article 5(1) of Council Regulation (EC) No 139/2004.[638] With the use of that notion and that reference, the Union legislature made clear that the ceiling for any fine imposed pursuant to Article 74 of Regulation (EU) 2022/2065 should take into account the financial capacity of the single economic unit providing the intermediary service in the Union and not just that of the legal entity of that unit operating in the Union. That argument also misrepresents the determination of the supervisory fee for X,[639] which follows a different methodology as set out in Commission Delegated Regulation (EU) 2023/1127. Not only can that methodology for determining the ceiling for the supervisory fee, set out in Article 43 of Regulation (EU) 2022/2065 not inform the calculation of the ceiling for the imposition of fines pursuant to Article 74 of Regulation (EU) 2022/2065, but the provider of X is also wrong to claim that it ceiling was determined on the basis of TIUC's consolidated net profit. Finally, contrary to the provider of X's claim, relying on the notion of 'undertaking' for calculating the 6% ceiling that applies to fines is not '*arbitrary, unfair and disproportionate*',[640] but mandated by Regulation (EU) 2022/2065.

(637) As explained in section 3 above, the Commission considers the provider of X to be the single economic unit ultimately controlled by Mr. Elon Musk. Since the notion of single economic unit is transposable to the notion of undertaking used in competition law and referenced in Article 3(x) of Regulation (EU) 2022/2065, the Commission considers that the relevant turnover for the purposes of calculating the 6% ceiling laid down in Article 74 of Regulation (EU) 2022/2065 is the total worldwide turnover of the single economic unit ultimately controlled by Mr. Elon Musk.

(638) In response to the sixth RFI, the provider of X only supplied information on the consolidated turnover of X Holdings Corp., which that provider considers to constitute the relevant undertaking for the purposes of Article 3(x) of Regulation (EU) 2022/2065. That turnover amounted to USD [redacted][641] (approx. EUR [redacted])[642] in 2024.

(639) Despite having been requested to do so, the provider of X did not provide any information on the total worldwide annual turnover for the single economic unit ultimately controlled by Mr. Elon Musk. Nor did it provide information on the total

635 *See* judgment of 13 February 2025, ILVA A/S, C-383/23, ECLI:EU:C:2025:84, paragraph 22.

636 Similarly, the Court of Justice has already clarified that for the specific context of the calculation of administrative fines imposed pursuant the General Data Protection Regulation, the concept of undertaking is to be understood within the meaning of Articles 101 and 102 TFEU. *See* judgment of 13 February 2025, ILVA A/S, C-383/23, ECLI:EU:C:2025:84, paragraph 21 and the case-law cited.

637 Reply to the Preliminary Findings, paragraph 365.

638 *See* judgment of 13 February 2025, ILVA A/S, C-383/23, ECLI:EU:C:2025:84, paragraph 36.

639 Article 63 of Regulation (EU) 2022/2065.

640 Reply to the Preliminary Findings, paragraph 364.

641 Reply to the sixth RFI, question 14.

642 ECB exchange rate for 2024 is on average USD 1 = EUR 0.9239.

worldwide annual turnover of X.AI Holdings Corp. for that matter. None of this information is in any event publicly available either.

(640) This absence of information should be considered when determining whether the amount of the fine remains within the limits of the maximum fine cap of 6 % of the total worldwide annual turnover for the single economic unit ultimately controlled by Mr. Elon Musk.

10.5. Amount of the fine

(641) In view of the criteria described in recitals 601 to 632 above, the Commission considers it appropriate to impose fines pursuant to Article 74(1), point (a) of Regulation (EU) 2022/2065 representing

– EUR 45 000 000 for the infringement of Article 25(1),

– EUR 35 000 000 for the infringement of Article 39, and

– EUR 40 000 000 for the infringement of Article 40(12) of Regulation (EU) 2022/2065.

(642) X Holdings Corp.'s total worldwide annual turnover in 2024, the financial year preceding the adoption of this Decision, was USD [redacted] [643] (approx. EUR [redacted])[644]. The total worldwide annual turnover of the single economic entity ultimately controlled by Mr. Elon Musk is likely to lay far above this amount. None of the individual fines imposed for the three infringements of Regulation (EU) 2022/2065 surpasses 6% of the total worldwide annual turnover of the single economic unit ultimately controlled by Mr. Elon Musk, which is the maximum fine that can be imposed pursuant to Article 74(1) of Regulation (EU) 2022/2065, let alone 6% of the total worldwide annual turnover of X Holdings Corp.

(643) The total level of the fine for each infringement is also proportionate, since it represents a relatively small proportion of the total worldwide annual turnover of the single economic unit ultimately controlled by Elon Musk, let alone X Holdings Corp.'s total worldwide annual turnover (approx. [redacted]% for the infringement of Article 25(1), [redacted]% for the infringement of Article 39 and [redacted]% for the infringement of Article 40(12) of Regulation (EU) 2022/2065). At the same time, the fine can still be considered sufficiently deterrent in relation to the single economic entity ultimately controlled by Mr. Elon Musk.

(644) In any event, the total amount of the three fines together, i.e. EUR 120 000 000 , does not exceed the 6% of the total worldwide annual turnover of the single economic unit ultimately controlled by Elon Musk or that of X Holdings Corp, even though the ceiling should be applied to the fine imposed for each of the three infringements individually.

11. PERIODIC PENALTY PAYMENTS

11.1. Principles

(645) Pursuant to Article 76(1), point e, of Regulation (EU) 2022/2065, the Commission may adopt a decision imposing on the provider of the very large online platform or of the very large online search engine concerned periodic penalty payments not

643 Reply to the sixth RFI, question 14.

644 ECB exchange rate for 2024 is on average USD 1 = EUR 0.9239.

exceeding 5% of their average daily income or worldwide annual turnover in the preceding financial year per day, calculated from the date appointed by that decision, in order to compel them to comply with a decision pursuant to Article 73(1) of that Regulation, including where applicable the requirements it contains relating to the action plan referred to in Article 75 of that Regulation.

(646) Pursuant to Article 75(4) of Regulation (EU) 2022/2065, when making use of the enhanced supervision system, the Commission may adopt periodic penalty payments where: (a) the provider of the very large online platform or of the very large online search engine concerned fails to provide any action plan, the audit report, the necessary updates or any additional information required, within the applicable period; (b) the Commission rejects the proposed action plan because it considers that the measures set out therein are insufficient to terminate or remedy the infringement; or (c) the Commission considers, on the basis of the audit report, any updates or additional information provided or any other relevant information available to it, that the implementation of the action plan is insufficient to terminate or remedy the infringement.

11.2. Application to this case

(647) In view of the seriousness of the provider of X's infringements of Articles 25(1), 39, and 40(12) of Regulation (EU) 2022/2065, as established in this Decision, and considering that the infringements of Articles 25(1), 39 and 40(12) of that Regulation have been found to be ongoing, the Commission concludes that it is necessary to impose periodic penalty payments pursuant to Article 76(1) of that Regulation on the provider of X, if that provider fails to: (i) implement measures that bring the infringement of Article 25(1) of that Regulation effectively to an end within 90 working days from the date of notification of this Decision; (ii) notify the Commission within 60 working days from the date of notification of this Decision of the specific measures by means of which the provider intends to bring the infringement of Article 25(1) of that Regulation effectively to an end and prevent the recurrence of that infringement; and (iii) draw up and communicate to the Digital Services Coordinators, the Commission, and the European Board for Digital Services within 90 working days from the date of notification of this Decision an action plan setting out the necessary measures which are sufficient to terminate or remedy the infringements of Articles 39 and 40(12) of that Regulation.

(648) Concerning the infringements of Articles 39 and 40(12) of Regulation (EU) 2022/2065, the Commission also concludes that it is necessary to impose periodic penalty payments on the provider of X, if that provider fails to take the steps listed in Article 75(4) of that Regulation relating to the action plan.

(649) The provider of X raises the same arguments against the imposition of periodic penalty payments as it does against the imposition of fines. However, Regulation (EU) 2022/2065 pursues objectives of great importance. Each day of continued non-compliance by the provider of X with Articles 25(1), 39 and 40(12) of Regulation (EU) 2022/2065 further compounds the resulting harm for society. Consequently, the imposition of periodic penalty payments is warranted.

(650) Any periodic penalty payment that may be definitively set should be sufficient to ensure compliance by the provider of X with this Decision and may take account of the provider of X's significant financial resources (see recital 637 of this Decision).

HAS ADOPTED THIS DECISION:

Article 1

The provider of X has not complied with Article 25(1) of Regulation (EU) 2022/2065 by deceptively designing, organising and operating X's online interface in relation to the 'verified' status feature of that service in a manner that materially distorts or impairs the ability of recipients of that service to make free and informed decisions regarding the authenticity and reliability of accounts on that service. Such non-compliance has commenced on 28 August 2023 and is ongoing.

Article 2

The provider of X has not complied with Article 39 of Regulation (EU) 2022/2065 by not providing an advertising repository, publicly available in a specific section of its online interface, through a searchable and reliable tool that allows multicriteria queries and through an application programming interface pursuant to Article 39(1) of Regulation (EU) 2022/2065 to obtain all the information on advertisements required by Article 39(2) of that Regulation for the entire period during which they present an advertisement and until one year after the advertisement was presented for the last time on its interface. Such non-compliance has commenced on 28 August 2023 and is ongoing.

Article 3

The provider of X has not complied with Article 40(12) of Regulation (EU) 2022/2065 by not giving access, without undue delay, to data that is publicly accessible on X's online interface to researchers qualifying for access pursuant to that provision. Such non-compliance has commenced on 28 August 2023 and is ongoing.

Article 4

For the infringement referred to in Article 1, a fine of EUR 45 000 000 is imposed on X Internet Unlimited Company jointly and severally with X Holdings Corp., X.AI Holdings Corp., and Mr. Elon Musk.

For the infringement referred to in Article 2, a fine of EUR 35 000 000 is imposed on X Internet Unlimited Company jointly and severally with X Holdings Corp., X.AI Holdings Corp., and Mr. Elon Musk.

For the infringement referred to in Article 3, a fine of EUR 40 000 000 is imposed on X Internet Unlimited Company jointly and severally with X Holdings Corp., X.AI Holdings Corp., and Mr. Elon Musk.

The fines shall be credited, in euros, within three months from the date of notification of this Decision, to the following bank account held in the name of the European Commission:

[redacted]

After the expiry of that period, interest shall automatically be payable at the interest rate applied by the European Central Bank to its main refinancing operations on the first day of the month in which this Decision is adopted, plus 3.5 percentage points.

Where an action pursuant to Article 263 of the Treaty is brought before the Court of Justice of the European Union against this Decision, the fines shall be covered by their due date, either by providing an acceptable financial guarantee or by making a provisional payment of the fines in accordance with Article 108 of Regulation (EU, Euratom) 2024/2509 of the European Parliament and of the Council.[645]

Article 5

The provider of X shall bring effectively to an end the infringements referred to in Articles 1, 2 and 3 in so far as it has not already done so.

The provider of X shall refrain from repeating any act or conduct described in Articles 1, 2 and 3, and from any act or conduct having the same or an equivalent effect.

Article 6

As regards the conduct described in Article 1, the provider of X shall (i) notify the Commission, within 60 working days from the date of notification of this Decision, of the specific measures through which it intends to comply with this Decision; and (ii) bring the infringement effectively to an end and prevent the recurrence of that infringement within 90 working days from the date of notification of this Decision.

Article 7

As regards the conducts described in Articles 2 and 3, the provider of X shall draw up an action plan setting out the necessary measures which are sufficient to terminate or remedy the infringements in accordance with Article 75(2) of Regulation (EU) 2022/2065.

The provider of X shall communicate the aforementioned action plan to the Digital Services Coordinators, the Commission, and the Board within 90 working days from the date of notification of this Decision.

Article 8

If the provider of X fails to comply with the orders set out in Articles 5, 6 or 7, it shall incur periodic penalty payments not exceeding the limit laid down in Article 76(1) of Regulation (EU) 2022/2065 from the date on which it is required to effectively comply with the orders set out in Articles 5, 6, or 7 until the date on which it complies with this Decision.

Article 9

This Decision is addressed to:

(a) X Internet Unlimited Company, One Cumberland Place, Fenian St, Dublin 2, D02 AX07, Ireland;

[645] Regulation (EU, Euratom) 2024/2509 of the European Parliament and of the Council of 23 September 2024 on the financial rules applicable to the general budget of the Union (OJ L 2024/2509, 26.9.2024).

(b) X Holdings Corp., 800 N STATE ST STE 304, DOVER, DE 19901-3925, United States of America;

(c) X.AI Holdings Corp., 1450 Page Mill Road, Palo Alto, California 94304, United States of America; and

(d) Mr. Elon Musk, as the natural person ultimately controlling the group of companies to which X Internet Unlimited Company, X Holdings Corp. and X.AI Holdings Corp. belong, Excession LLC, 2110 Ranch Road 620 South, #341886, Austin TX 78734, United States of America.

This Decision shall be enforceable pursuant to Article 299 of the TFEU.

Done at Brussels, 5.12.2025

For the Commission

Henna Virkkunen
Executive Vice-President

CERTIFIED COPY
For the Secretary-General

Martine DEPREZ
Director
Decision-making & Collegiality
EUROPEAN COMMISSION