



DSA RISK ASSESSMENT REPORT 2023

Confidential

29 September 2023

CONTENTS

1. FOREWORD	2
2. INTRODUCTION	3
3. KEY INFORMATION ABOUT TIKTOK	3
4. EXECUTIVE SUMMARY	6
5. RISKS OF CHILD SEXUAL ABUSE MATERIAL AND CHILD SEXUAL EXPLOITATION	9
Risk Mitigations - Table 1: Child sexual abuse material and child sexual exploitation	12
6. ONLINE PROTECTION OF MINORS AND ASSOCIATED RISKS	16
Risk Mitigations - Table 2: Online protection of minors and associated risks	19
Case Study: Online challenges and hoaxes	25
Deep Dive: Content levels for Younger Users	27
7. RISKS TO ELECTIONS AND CIVIC INTEGRITY	28
Risk Mitigations - Table 3: Risks to elections and civic integrity	31
Case Study: Spain's parliamentary election (July 2023)	36
8. RISKS OF GENDER-BASED VIOLENCE CONTENT	37
Risk Mitigations - Table 4: Risks of gender-based violence content	40
9. RISKS OF TERRORIST CONTENT	43
Risk Mitigations - Table 5: Risks of terrorist content	46
Case Study: Proactive mitigation measures relating to violent extremism risks	49
10. RISKS OF ILLEGAL HATE SPEECH CONTENT	51
Risk Mitigations - Table 6: Risks of illegal hate speech	54
11. RISKS TO PUBLIC HEALTH FROM MEDICAL MISINFORMATION CONTENT	58
Risk Mitigations - Table 7: Risks to public health from medical misinformation	61
12. RISKS TO PUBLIC SECURITY FROM HARMFUL MISINFORMATION/CONTENT	64
Risk Mitigations - Table 8: Risks to public security from harmful misinformation/content	67
Case Study: Handling civil unrest in France	70
13. RISKS TO FUNDAMENTAL RIGHTS	71
Risk Mitigations - Table 9: Risks to fundamental rights	74
Deep Dive: Striking a balance between preventing harm and enabling expression	77
14. RISKS OF INTELLECTUAL PROPERTY INFRINGING CONTENT	79
Risk Mitigations - Table 10: Risks of intellectual property infringing content	81
ANNEX 1 - RISK ASSESSMENT METHODOLOGY	84
ANNEX 2 - HOW TO USE THIS REPORT	85

1. FOREWORD

TikTok's mission is to inspire creativity and bring joy. Every day millions of Europeans¹ come to TikTok to find entertainment, education and have fun. TikTok is a place where creativity thrives, a space where anyone and everyone can express themselves, share their passion or even their profession. TikTok is a place to [learn](#), to [celebrate music](#), and a place for [cultural discovery and creative self-expression](#).

Our teams continually work to make TikTok a place where everyone can express their creativity and enjoy a wide range of content. Thousands of Trust & Safety professionals are focused every day on helping to make our Platform safe and welcoming for our community. Much of this work is led by our EMEA Trust & Safety teams in Dublin.

The Trust and Safety teams at TikTok are focused on carrying out a wide variety of tasks to protect our community from a range of risks, and many of these safety measures are set out in this Report. The challenges faced by online platforms like ours are complex and constantly evolving, which makes it vitally important that we invest in building internal teams that have deep subject matter expertise across their respective fields, from hate speech to violent extremism, and election misinformation to minor safety. We complement the work of our expert teams through close collaboration with our European regional and country-level teams, ensuring that TikTok brings a combination of both subject matter expertise and local insight to our safety work.

The complex nature of these challenges also makes it vital to have constructive engagement with a wide range of outside experts and stakeholders, including through our [Safety Advisory Council](#) for Europe and through our plans to establish [TikTok's Youth Council](#) later in 2023. From youth safety to hate speech, these forums provide our teams with crucial opportunities to hear from outside experts and to listen to the experiences of those who directly use our Platform. They also act as critical sounding boards as we continue to develop new ways to promote safety on and off TikTok. We look forward to further engaging with research and civil society communities.

The work of our Trust & Safety teams is reflected in our quarterly [Community Guidelines Enforcement Reports](#): over a 12-month period from April 2022 to March 2023, our teams and technology have proactively removed 96.3% of identified violative video content before it was reported to TikTok; removed 91.9% of violative video content within 24-hours of it being posted; and removed 86.4% of violative videos before they received any views. We are proud of these numbers, but constantly strive to do more.²

We therefore welcome the DSA, as we see it as a way to further help us deliver on our mission to inspire creativity and bring joy, and to keep our Platform safe. We welcome the enhanced transparency that the DSA brings and the important role of risk assessments, as we see these as ways to hold ourselves to account and to continuously improve in managing risks.

¹ "Europeans" refers to TikTok users located in countries of the European Union.

² 'Proactive removal' when used to refer to data around content removals means: identifying and removing a video before it's reported to TikTok by any means. Removal within 24 hours means removing the video within 24 hours of it being posted on the Platform.

While we have made important progress up to Day 1 of the DSA, we do not see a 'finish line' when it comes to safety. It is our full intention to continue to assess our current policies and processes and listen to and take on board candid feedback. We believe the important work undertaken in preparation for Day 1 of the DSA has helped put us in a strong position to tackle today's challenges, to prepare for tomorrow's threats, and ultimately to keep bringing joy, entertainment, and connection to people across Europe and around the world.

Cormac Keenan
Director, TikTok Technology Limited
Global Head of Trust and Safety, TikTok

2. INTRODUCTION

This DSA Risk Assessment and Mitigation Report (the "**Report**") has been prepared in accordance with Art. 42(4) of the EU Digital Services Act (Regulation 2022/2065) ("**DSA**"). It has been prepared by TikTok Technology Limited ("**TikTok Ireland**") in relation to the operation in the European Union ("**Europe**") 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. This Report has been reviewed and approved by the board of directors of TikTok Ireland, following consultation with TikTok's head of the compliance function.

3. 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 has 134 million monthly active users in Europe³.

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. TikTok's features include video, photo, livestream and comments. Users can share, skip, swipe, like, comment on, or replay videos.⁴ 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.

³ <https://www.tiktok.com/transparency/en/eu-mau-2023-7/>

⁴ Whether an account is private (meaning only people a user approves can follow them, view their profile, and watch their videos) or public (everyone can choose to follow the user, watch their videos and view their profile), users can limit the audience for their videos whenever they post a video.

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.⁵ 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. TikTok operates a range of processes in order to prevent harmful content from appearing in the FYF and provides an additional layer of controls for minors.

What does advertising look like on TikTok?

Businesses show advertising (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 to being transparent with its users about how it collects, uses and shares data for ads. Minors do not see ads based on profiling but will see generic ads.

TikTok’s Ad 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 very much in 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. Users can also report the ad if they consider it to be violative of TikTok’s Ad Policies.

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 Ad 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 to it by users or third parties.

TikTok operates its content moderation processes using automated and manual (human) means in accordance with the following 4 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 it is only viewed by adults (18 years and older);
3. Maintain FYF eligibility standards to help ensure any content that may be promoted by its recommendation system is appropriate for a broad audience; and
4. Empower its community with information, tools, and resources.”⁶

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.

⁵ <https://support.tiktok.com/en/using-tiktok/exploring-videos/how-tiktok-recommends-content>

⁶ <https://www.tiktok.com/community-guidelines/en/>

All video, photo and text-based content⁷ uploaded to the Platform firstly goes through 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 Ad Policies (as applicable).

TikTok's moderation 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.

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 actions up to, and including, a permanent account ban. 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 adopts a range of processes to review the accuracy of decisions when moderating potentially violative content. TikTok's policies and processes for detecting and removing violative ads are similar to those set out above.

As of TikTok's DSA Day 1, users can also report user generated and ad content which they consider to be illegal under European or Member State law. In addition, TikTok's European Online Safety Hub⁸ contains guidance on how to use that reporting function, which includes reporting illegal content in the categories contained in this Report.

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, a backup methodology and forensic capabilities to ensure the redundancy of the TikTok infrastructure, their recovery in case of an incident and the timely restoration of data. 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, UK 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 objection, restriction, deletion, rectification and portability of their personal data.

⁷ In accordance with Recital 14 of the DSA, interpersonal communication services (as defined in Directive (EU) 2018/1972 and to the extent provided on the Platform) are excluded from this description.

⁸ See <https://www.tiktok.com/euonlinesafety/en/>

4. EXECUTIVE SUMMARY

Introduction

- This Report summarises the results of TikTok's first systemic risk assessments and the specific mitigation measures in place under Arts. 34 and 35 of the DSA. This Report has been prepared solely for this purpose. Please see 'How to use this report' for more information.
- TikTok's focus in performing these risk assessments has been to:
 - Define systemic risk areas and contextualise them in Platform design, functioning, and use;
 - Consider the severity and probability of these risks, taking account of mitigations that TikTok has implemented;
 - Where appropriate, to identify further mitigation actions that are reasonable, proportionate, and capable of reducing the level of risk; and
 - To reach conclusions about priority risk areas, which will inform TikTok's ongoing risk management strategy.
- TikTok has dedicated substantial resources and time to developing a risk assessment process designed to meet the specific requirements set out under the DSA. This work has been led by a core team, with multiple subject matter expert groups focused on specific risks.
- A summary of TikTok's risk assessment methodology is set out in Annex 1 to this Report.

Risk Environment

- TikTok is one of a number of online platforms which face similar risks. These risks often transcend the boundaries of individual platforms and involve complex interactions between online experiences and real-world events.
- Like all online platforms, the Platform facilitates expression and the receiving and imparting of information and ideas. This is a central value of the European Union, a fundamental right, and a core pillar of European democracy and social life.
- The Platform hosts billions of items of content and recommends content to its users depending on signals about what users may want to see. TikTok strives to only host content within the bounds of the rules TikTok imposes on itself, its users, and relevant laws.
- This necessarily involves the careful balancing of risks related to fundamental rights and user safety, both online and in the real world. Operating within such complexity and considering the inherent risks of human expression, it is not possible to prevent all risks. As the Commission describes, there will always be content that constitutes potential risk.⁹ TikTok strives to proactively identify and remove content before it is seen, while reporting on its content removal volumes, the proportion proactively detected and removed, the proportion removed before any user views it, and the proportion removed within 24 hours¹⁰.
- TikTok's own rules, which it enforces through its content moderation policies and processes, reflect an approach that prioritises caution in areas where risks are more likely to impact real world harms and minors. TikTok has implemented a combination of automated content moderation and human reviews to detect and remove harmful content promptly (see 'Key Information about TikTok' above).

⁹ See the report of the European Commission: Application of the Risk Management Framework to Russian disinformation campaigns, page 15.

¹⁰ See TikTok's quarterly Community Guidelines Enforcement Reports
<https://www.tiktok.com/transparency/en-us/community-guidelines-enforcement-2023-1/>

- TikTok further considers potential systemic risks by taking into account its wide appeal across demographics, regions, and cultures. The Platform's primary function is the sharing of audiovisual content by one user to many, rather than functioning as a platform which is focused on one to one communications. Whilst TikTok is not specifically aimed at minors or predominantly used by them, it places a particular focus on how risks may manifest for and impact those users. TikTok has considered the physical and mental well being of both minors and all users in the risk assessments that inform this Report.
- Finally, 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.

Risk assessment results

- TikTok applies considerable resources and mitigations to addressing all the systemic risks considered in this Report. TikTok has taken a cautious approach to assessing the severity and probability of possible systemic risks that may stem from the Platform or the use made of it. In order to demonstrate the resulting priority areas for TikTok, it has organised the results of the risk assessments into the tiers set out below.
- These tiers represent TikTok's current assessment of the priority that a systemic risk category demands, having taken account of existing policies, systems and procedures for mitigating the risk. TikTok has also given weight to risks impacting vulnerable groups or which could be most likely to arise (potentially with high velocity) in the next 12 months.
- Moving forward, these tiers will inform TikTok's view of when further mitigations are needed, and what kind of mitigation would be reasonable, proportionate, and effective. Tiers will be reviewed and updated, depending on the results of ongoing systemic risk assessments.

SUMMARY OF RISK ASSESSMENT RESULTS

Tier 1 risks (sections 5 to 8 of this Report)	Tier 2 risks (sections 9 to 12 of this Report)	Tier 3 risks (sections 13 to 14 of this Report)
<ul style="list-style-type: none"> • Risks of child sexual abuse material and child sexual exploitation • Online Protection of Minors and associated risks • Risks to elections and civic integrity • Risks of gender-based violence content 	<ul style="list-style-type: none"> • Risks of terrorist content • Risks of illegal hate speech content • Risks to public health from medical misinformation • Risks to public security from harmful misinformation/content 	<ul style="list-style-type: none"> • Risks to fundamental rights • Risks of intellectual property infringing content
<ul style="list-style-type: none"> • TikTok's assessment mapped existing mitigation policies, processes, and systems to areas where it was identified that systemic risk could arise. These mitigations have been 		

categorised according to relevant mitigation measures outlined in DSA Art. 35(1). In respect of each risk, the most significant mitigations are described in each section of the report below.

- TikTok has also identified a range of further mitigation improvements. These mitigations are summarised in each section of the report below.

Emerging risks

- TikTok is alert to the emerging risks of synthetic media, also known as generative AI. This technology makes it increasingly easy to create realistic images, video, and audio that can make it more difficult to distinguish between fact and fiction and the creation of illegal content, or can be used as part of image-based abuse.
- 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.
- TikTok has also taken proactive measures to mitigate generative AI risks in a reasonable and proportionate way:
 - TikTok has created a policy prohibiting synthetic media showing realistic scenes that are not prominently disclosed or labelled in a video;
 - TikTok does not allow any synthetic content of a real adult (with the exception of adults with a public profile) without that being disclosed and does not allow depiction of any real child, even if it contains non-violative content;
 - Creators have the ability to label their content "Creator labelled as AI-generated" where it is either completely generated by AI, or significantly edited. Creators are required to do so if the content appears realistic and could mislead users; and
 - Users are able to report content where they believe that it is undisclosed synthetic media or where synthetic media otherwise violates the Community Guidelines. For example, synthetic media that contains the likeness of any private natural person.
- 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.

Next steps for TikTok

- This first Report necessarily forms a baseline which is based on extensive information collated and analysed before the start of the first year of the DSA coming into force. TikTok will continue to mature and improve its approach to the identification and mitigation of systemic risks where these can stem from the design or functioning of the Platform and its related systems, or the use made of the Platform.
- TikTok will monitor the delivery of the mitigation developments described in this report, and their impact. TikTok will also identify critical updates to its service and ensure that the risk assessments remain up to date. This will necessarily involve stakeholders at all levels of the business and TikTok is developing its enterprise risk management system to achieve this.
- Looking ahead, TikTok welcomes feedback on this Report from its regulators (the European Commission and TikTok's Digital Service Coordinator of Establishment, Ireland's Coimisiún na Meán) and from the research and civil society stakeholders that will engage with TikTok via this Report or otherwise.

5. RISKS OF CHILD SEXUAL ABUSE MATERIAL AND CHILD SEXUAL EXPLOITATION

Description of the risk:

- TikTok understands the terms child sexual abuse material (“**CSAM**”)¹¹ and child sexual exploitation (“**CSE**”) in a manner consistent with EU and EU member state laws. The scope of the CSAM and CSE risks have been defined in particular having regard to Directive 2011/93/EU, Arts. 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 off Platform to obtain or distribute CSAM content;
 - 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/avatar and profile bio that contains CSAM content.
- TikTok considers that the activities associated with CSE 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/avatar and profile bio that involves CSE behaviour.
- In addition, TikTok notes that Art. 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 Art. 24 of The Charter of Fundamental Rights of the European Union (the “**Charter**”). TikTok further 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

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 \(known as the “Luxembourg Guidelines”\)](#), we consider “CSAM” to be the more appropriate term to be used throughout this Risk Assessment.

¹² [UN General Comment No. 25 \(2021\)](#) on Children’s Rights in Relation to the Digital Environment.

Key mitigation measures put in place:

- Risk Mitigations - Table 1 sets out a summary of risk mitigation measures in place with reference to Art. 35 (1) (a) to (k) of the DSA.
- Key mitigations for this risk are: (1) the adaptations that TikTok has made to its content moderation systems and processes to specifically detect CSAM and CSE content; (2) third party collaboration to make sure that detection models are trained using on and off Platform data; and (3) the specialist Child Safety Team that are dedicated to the detection and removal of CSAM and CSE content (who together with other relevant teams ensure that local and linguistic factors are identified and addressed); and (4) limitations on communications between minors and adults on the Platform and the proactive review of all video, photo and livestream content, even if posted to private accounts.

Key data relied on:

- The National Center for Missing and Exploited Children (“**NCMEC**”)’s CyberTipline is a centralised reporting system for the online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement. In [2022](#), on a global basis (including data from the EU) NCMEC received 288,125 reports of suspected CSAM or CSE from TikTok. This was an increase from the approximately 154,618 reports received from TikTok in [2021](#). TikTok reports made up approximately 0.9% of all reports submitted by participating platforms in 2022.
- TikTok’s Community Guidelines reporting categories (under the more general category of “Minor Safety”), as reflected in the Q1 Community Guidelines report, do not contain a separate report on CSAM/CSE. However, in relation to all content removed globally for Minor Safety reasons, 98.9% of removals were made proactively, 88% of which were made before any user viewed them. 91.2% of content was removed within 24 hours of being uploaded.

Severity:

- Dissemination of CSAM and CSE risks are among the most severe risks that can arise on an online platform. Such risks can involve a real-world risk of serious harm to the physical and psychological safety of victims, as well as the risk of having severe psychological impact on those who may be inadvertently exposed to such content.
- In terms of duration/remediability, the impacts of the dissemination of CSAM or CSE are likely to have long-term impacts on victims and it is likely very difficult to restore the prevailing situation prior to such risks arising if a child has been physically harmed or exploited. The effects are likely to be less severe for those inadvertently exposed to such content.
- TikTok’s key data above demonstrates that TikTok’s mitigation measures have a significant impact on the scale and potential duration of harm to users from CSAM and CSE.
- TikTok assesses the potential severity of the risk of any dissemination of CSAM and CSE to be material,¹³ due to: (1) the nature of harm involving a risk to the physical and psychological state of victims and the impact being long term and likely very difficult to reverse; and (2) TikTok’s proactive efforts to enforce its policies, as demonstrated by the mitigation measures referred to above and in particular TikTok’s 98.9% proactive detection rate (of all minor safety related violations) referred to above. In particular, this assessment results from the measures put in place to detect and prevent the dissemination of CSAM or CSE at scale, but still noting

¹³ Following the update to TikTok’s Community Guidelines in March 2023, we expect that future reports will better reflect this risk category under the policy section entitled “Youth Exploitation and Abuse”.

the severe nature of CSAM and CSE and of the potential long-term impacts on victims.

Probability:

- CSAM is manifestly illegal content and generally identifiable on its face, and therefore not appealing to broad audiences. TikTok's functionality makes it hard for users to discover specific types of material intentionally. TikTok therefore concludes that a bad actor is less likely to choose to upload photo and video content to TikTok for broad public consumption, in preference over other more limited means of clandestine communication that are available. This conclusion is supported by independent research, which considered the issue in the specific context of self-generated CSAM.¹⁴
- Given that it is highly criminal in nature, CSAM content is very unlikely to be popular among audiences, such that it is selected to be recommended for users to view via the For You page, or have the potential to be widely disseminated on TikTok. For the same reason, it is also unlikely to be included in any ads content on the Platform.
- TikTok strives to prevent the upload of, or otherwise remove, CSAM and CSE content from the Platform. TikTok assesses that, due to the inherent nature of the Platform and the limited volumes that have been identified in the past, it is unlikely that there is widespread dissemination of CSAM or CSE content on the Platform.

Key stakeholder engagement:

- TikTok's European Law Enforcement Outreach team engages with Europol and specialist law enforcement authorities across the EU and internationally. In particular, this involves engagement with specialist units focused on child safety and counter-terrorism matters.
- TikTok works with a range of civil society groups and has ongoing partnerships with leading online safety organisations, including: the Technology Coalition, NCMEC, the Internet Watch Foundation, Family Online Safety Institute, ConnectSafely, INHOPE Association, WePROTECT Global Alliance, DQ Institute, Thorn, and the Lucy Faithfull Foundation.

Prioritisation:

- TikTok considers risks of dissemination of CSAM and CSE content to be a **Tier 1** priority.
- This is due to the severity of the possible risk of physical and psychological harm to minors and users exposed to such egregiously illegal content. TikTok takes this view notwithstanding its analysis that TikTok's functionality does not lend itself to the amplification of these risks.
- TikTok's specialist teams will continue to closely monitor and remain highly vigilant of these sensitive risks. TikTok will invest in further improvements in its mitigations and will continue to enforce its Community Guidelines policies on [Youth Safety and Well-being](#) and [Youth Exploitation and Abuse](#), and continue to dedicate specialist CSAM and CSE prevention resources to detect and mitigate such risks.


Key further mitigation effectiveness improvements in line with Art. 35 of the DSA:

- **Detection developments (Art. 35(1)(f)):** TikTok will look to further expand its proactive detection using hashing technology to additional product features. [REDACTED]

¹⁴ See Stanford Internet Observatory Cyber Policy Centre study dated 7 June 2023, <https://stacks.stanford.edu/file/druid:jd797tp7663/20230606-sio-sg-csam-report.pdf>

- **External engagement (Art. 35 (1)(g)):** TikTok will engage with the organisations newly designated as “trusted flaggers” by EU member states (per Art. 22 DSA), and undertake outreach to onboard such entities to ensure efficient and priority processing of such reports via TikTok’s dedicated channels.
- **Continued risk monitoring and vigilance (Art. 35(1)(f)):** TikTok will continue to devote cross-functional resources on a priority basis to its processes for handling CSAM and CSE-related risks. In addition, TikTok shall continue to collect and monitor relevant data as part of its transparency reporting obligations under the DSA.

Risk Mitigations - Table 1	Child sexual abuse material and child sexual exploitation
TikTok’s risk-mitigation measures in accordance with Art. 35(1) of the DSA (a) to (k), plus any other relevant measures:	
(a) Adaptation of feature or platform design, including online interfaces (as defined in Art. 3(m) DSA):	<p>TikTok has made a number of adaptations to its platform to mitigate the risk of CSAM content and CSE behaviour, including (but not limited to) the following:</p> <ul style="list-style-type: none"> • TikTok has implemented various settings for 13-17 year olds to limit who can view their content and interact with them on the Platform; • TikTok has implemented a range of targeted measures to address the risk of CSAM being shared and/or accessed in post-to-private (i.e. whereby one attempts to circumvent safety measures by posting CSAM content in a private account and then share the log-in details with others) [REDACTED]; • TikTok may, from time to time, block [REDACTED] from appearing on the Platform; and • Only users aged 18 and over can host livestreams. Livestreams are subject to similar automated review processes that are tailored to take account of the particularities of the livestream environment, including a review when a certain threshold of viewers is reached. If a minor safety-related violation is detected in livestreams, the livestream is shut down.
(b) Adaption of terms and conditions (as defined in Art. 3(u) DSA) and their enforcement:	<p>TikTok very clearly prohibits CSAM and CSE on the Platform through a combination of TikTok’s Terms of Service and the Community Guidelines, and takes a range of measures to generate awareness and inform users.</p> <p>The Terms of Service expressly incorporate TikTok’s Community Guidelines, which state ‘We do not allow youth exploitation and abuse, including child sexual abuse material (CSAM), nudity, grooming, sextortion, solicitation, pedophilia, and physical or psychological abuse of young people’ and which provides detailed further explanation.</p>

	Independent research, undertaken by the Stanford Internet Observatory Cyber Policy Centre, indicates that TikTok has the strictest policy approach to CSAM and CSE-related content/behaviour among peer platforms. ¹⁵
(c) Adoption of content moderation (as defined in Art. 3(t) DSA) processes:	<p>Please see the 'Key Information about TikTok' section for a description of TikTok's content moderation processes.</p> 
(d) Testing and adaption of algorithmic systems, including recommender systems (as defined in Art. 3(s) DSA):	<p>TikTok takes a range of measures to ensure that manifestly illegal content such as CSAM content is removed prior to publication or is detected and removed from the Platform as soon as possible. By extension, those measures are also designed to ensure that such content is not recommended or surfaced in a user's FYF.</p> <p>TikTok has protocols in place to quickly detect and control trends featuring harmful user content, such as CSAM content. Potential measures that can be deployed as part of its control strategies include banning certain hashtags and/or blocking searches associated with emerging risks.</p>
(e) Adaptation of advertising systems:	Please see the 'Key Information about TikTok' section for a description of TikTok's content moderation processes.
(f) Reinforcing risk detection measures:	<p>TikTok operates a number of specialist teams that deal with illegal content and CSAM and CSE content in particular and who play a role in reinforcing risk detection measures. These include, but are not limited to:</p> <ul style="list-style-type: none"> • TikTok's Child Safety Team provides around-the-clock 24/7 coverage, and includes a number of experienced child safety specialists who have accrued considerable investigatory and child safeguarding experience within the technology industry. The team is responsible for handling escalation of suspected CSAM content and CSE and for making reports made to NCMEC and to law enforcement authorities. • TikTok's Trust & Safety Risk Analysis team comprises a designated Risk Analysis team staffed by experienced professionals with backgrounds in cyber intelligence and risk detection. This team's activities include monitoring open-source resources and reporting to cross-functional colleagues on potential emerging risks. • TikTok has a dedicated Law Enforcement Response Team with responsibility for reviewing the legal validity of requests for user data from law enforcement authorities and government agencies across Europe. In addition, TikTok's dedicated Law Enforcement Outreach team engages in outreach with national law enforcement authorities across Europe and with agencies such as Europol on the investigation of child safety matters. This outreach supports TikTok's risk detection and intelligence gathering on new and emerging CSAM and CSE risks.

¹⁵ Cross-Platform Dynamics of Self-Generated CSAM, David Thiel, Renée DiResta and Alex Stamos, Stanford Internet Observatory (Cyber Policy Centre) v1.2.0 (2023-06-07), page 10.

<p>(g) Cooperation with trusted flaggers:</p>	<p>TikTok already operates a Community Partner Channel, so that onboarded NGOs perform a similar role to designated trusted flaggers (under the DSA) to submit reports of suspected harmful content. These NGOs can then report suspected CSAM content directly to TikTok's Trust & Safety teams, who review such reports on a priority basis. Those partners include the Italian Garante, Pharos (France's cyber police agency), Child Focus (Belgium), the Internet Watch Foundation, Portugal's Safer Internet Helpline and Save the Children.</p>
<p>(h) Cooperation with other platforms through the codes of conduct/crisis protocols:</p>	<p>TikTok notes that no codes of conduct under Art. 45 of the DSA or any crisis protocols under Art. 48 of the DSA have been adopted. However, TikTok does cooperate with other platforms in a number of ways to combat CSAM and CSE, as TikTok has integrated the following tools into its mitigation measures:</p> <ul style="list-style-type: none"> • TikTok uses the Internet Watch Foundation's hash lists to detect, remove and report known CSAM; • TikTok uses the NCMEC's flagship hash sharing platform and electronic service provider industry sharing hash lists. These programs enable TikTok to detect, remove and report known CSAM and share hashed values to further combat the spread of CSAM across participating platforms; • CSAI Match from YouTube, which is an API solution that helps identify re-uploads of previously identified CSAM content in videos; and • Content Safety API from Google, which assists in proactive detection of not-before-seen CSAM imagery so it can be reviewed and, if confirmed as CSAM, removed and reported.
<p>(i) Awareness-raising measures for recipients of the services:</p>	<p>TikTok adopts a range of measures to enhance media literacy and generate awareness of risks relating to CSAM and CSE and the resources and tools that are available (for example, instructions on how to report such content) as part of its risk mitigation strategy in relation to CSAM and CSE.</p> <p>This is a multifaceted approach involving integrated measures, which includes (but is not only) TikTok's Safety Centre, which includes a page on Preventing child sexual abuse on TikTok. Users are directed to this page if/when they search for CSAM-related terms. This initiative was developed following consultation with the NGO, InHope. This includes instructions on "How to report sexual content of someone under 18".</p>
<p>(j) Targeted measures to protect the rights of the child:</p>	<p>By their nature all measures that mitigate risks of CSAM and CSE are measures to protect the rights of the child. These are set out in more detail in the 'Online Protection of Minors' section of this Report, where all minor safety issues are reported upon.</p>
<p>(k) Measures to identify and address inauthentic content and behaviours:</p>	<p>As reflected in TikTok's Community Guidelines Enforcement Reports (for Q1 2023), fake engagement, spam account activity and covert influence operations are real threats that exist on the Platform.</p> <p>Inauthentic accounts can be used to facilitate off Platform sale and trade of CSAM material and/or may be used to pose as minors or other vulnerable individuals to manipulate and groom potential victims. In relation to CSE,</p>

these accounts can be used to establish trust and exploit their targets over time. TikTok's policies prohibit impersonation including leveraging TikTok accounts under false or fraudulent pretences. [REDACTED]

[REDACTED] Once detected, these accounts and the behaviours associated with them (such as [REDACTED]) are removed from the Platform. TikTok also manually monitors user reports of inauthentic accounts in order to detect larger clusters or similar inauthentic behaviours. TikTok's Community Guidelines also make clear that CSAM and CSE content that is both real and also fictional or digitally created is prohibited.

6. ONLINE PROTECTION OF MINORS AND ASSOCIATED RISKS

Description of the risk:

- TikTok is a Platform accessible to minors and therefore gives careful and appropriate consideration to the risk encompassed by Art. 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 acknowledges that a balance must be found between the design of the measures to address risks to minors and the fundamental rights of legitimate users of any age. Those rights include the right to freedom of expression, which includes the right 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 minor safety measures summarised in this Report.
- TikTok recognises the following minor safety risks:
 - **“Mis-stated Age Risk”**: TikTok’s minimum age requirement is 13 years old. By mis-stating their age at registration, minors aged under 13 (under the Platform’s minimum age requirement) may attempt to gain access to the Platform (**“Underage Users”**), and minor users aged 13-17 (**“Younger Users”**) may not receive an age-appropriate experience on the Platform;
 - **‘Content Risk’**: Minors 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’**: In creating and posting content on the Platform, or by engaging with content posted by others, minors 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.¹⁶ 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).

Key mitigation measures put in place:

- Risk Mitigations - Table 2 sets out a summary of risk mitigation measures in place with reference to Art. 35 (1) (a) to (k) of the DSA.
- Key mitigations include: (1) operating a neutral age gate and proactive processes to detect Underage Users; (2) limiting access to certain product features depending on age; (3) developing Content Levels that sort content by levels of thematic maturity (see Deep Dive “Content levels for Younger Users” below); (4) using restrictive default privacy settings; and (5) making content created by anyone under 16 ineligible for the FYF.
- TikTok is not specifically aimed at or predominantly used by minors aged under 18 in Europe. TikTok’s mitigation measures require an appropriate and proportionate balancing of the risks relating to minor safety and in particular the Mis-stated Age Risk, given the potential

¹⁶ CSAM and CSE risks are dealt with in the above section of this Report on risks of child sexual abuse material and child sexual exploitation.

impact that those measures may have on the fundamental rights of minors and adults, including freedom of expression and privacy. TikTok aims to strike this balance taking leading industry practice into account. TikTok considers that the technical restrictions imposed on and/or offered to minors and their parents/guardians on its Platform are proportionate and in the best interests of Younger Users and provide those users with an appropriate level of autonomy and control over their experience on TikTok.

Key data relied on:

- In relation to the Mis-Stated Age Risk, TikTok publishes the volume of underage user accounts aged under 13 that it has detected and removed from the Platform on a quarterly basis: [Community Guidelines Enforcement Report \(Q1 2023\)](#). For the period January to March 2023, the number of removals for the EEA was 2,203,894.
- In relation to the Content, Conduct and Contact Risks, TikTok reported in its Q1 2023 Community Guidelines Enforcement Report that it detected and removed 28,504,537 pieces of video content globally under TikTok's policies and processes for minor safety¹⁷. In addition:
 - 98.9% of such content was detected and removed proactively, before any reporting;
 - 88% was removed before there were any views of that content; and
 - 91.2% was removed within 24 hours of upload.

Severity:

- Minor safety is a top priority for TikTok. The precise severity of the Mis-stated Age, Content, Conduct and Contact risks is challenging to quantify as it will depend on several factors, including but not only, the age of the user, the specific user's personal maturity and other characteristics, their interactions with the app, the nature of content that they see and the amount of time they are exposed to content that is not age-appropriate.
- The enforcement of the Community Guidelines through proactive content moderation measures and extensive safety features provides a foundational set of mitigations before the application of specific minor safety mitigations.
- TikTok's Community Guidelines prohibit some mature content which may be available on other platforms (e.g., nudity, sexual activity). This deliberate approach takes into account that minors may be exposed to themes and content which may not always be suitable. These measures serve to materially reduce the potential scale of minor's exposure to harmful content.
- The Mis-stated Age Risk may also be of a short term duration, given that Underage User accounts are deleted from the Platform upon detection, so that such risks are capable of remediation. TikTok acknowledges that it is a challenge to assess the scale of the risk.
- The Content, Conduct and Contact Risks will last for as long as a user of the Platform is a minor and can be remediated via a series of interventions regarding the types of content on the Platform and how Younger Users can interact with them, by placing proportionate controls on user interactions, together with user education measures.
- TikTok assesses the potential severity of these risks in connection with the online protection of minors to be moderate due to: (1) the nature and duration of the harms, which vary in potential impact and duration but nevertheless impact TikTok's more vulnerable users; and (2) TikTok's proactive efforts to enforce its policies, as demonstrated by TikTok's 98.9% proactive detection rate for harmful content and the volume of underage user accounts removed. In particular this

¹⁷ TikTok considers its video removals under the policies contained within 'minor safety' to be generally indicative of its content removal data in relation to the Content Risk. More broadly, see TikTok's Community Guidelines Enforcement Reports, in particular in relation to those policies targeting Bullying and Harassment, Suicide and Self harm and Dangerous Acts and Challenges.

assessment results from the measures put in place to reduce negative impacts through specific adaptations to the functionality of the Platform.

Probability:

- An assessment of the risks relating to the online protection of minors involves considering and balancing a number of different elements and TikTok considers that there is no single indicator that demonstrates the probability of harm.
- The extent of TikTok's detection and removal of Underage User accounts - through proactive moderation efforts and reporting mechanisms - is an indicator of the scale of the Mis-stated Age Risk as regards Underage Users.
- The video removal rates referred to above are relative to the strict parameters of TikTok's Community Guidelines, which prohibit a wide range of content, from minors in minimal clothing in any context, to more harmful content including CSAM and CSE.
- TikTok strives to provide appropriate and proportionate measures to ensure a high level of privacy, safety and security for minors on its Platform. However, as with any online platform of a similar nature to the Platform, it is possible that minors may access content that is not age appropriate or engage in contact with other users that is not age appropriate, bearing in mind their age and level of maturity. Further improvements in age assurance technology (acknowledging that there is no established state of the art) and content moderation practices would reduce the likelihood further.

Key stakeholder engagement:

- TikTok has established [Safety Advisory Councils](#) which are an important source of advice;
- TikTok consults with various industry bodies, NGOs and external experts who provide information and insights that simultaneously inform and confirm risks to Minor Safety. These include, but are not only: FOSI, [WePROTECT Global Alliance](#), the Cyberbullying Research Center, The Safer Internet Center (Italy) and the Verification of Children Online project;
- TikTok also has ongoing partnerships with leading online safety organisations, including ConnectSafely, [INHOPE Association](#), [DQ Institute](#), Thorn, and the Lucy Faithfull Foundation;
- TikTok consults with academic research and experts from the [Digital Wellness Lab](#) at Boston Children's Hospital in connection with various Platform design aspects, including its [screen time](#) management tools (which set a default 60-minute limit for teens and provides several different options for Family Pairing accounts); and
- As part of an international research project into online challenges, TikTok has worked with academic and minor safety experts, including Dr. Zoe Hilton, Praesidio Safeguarding, and Western Sydney University Young and Resilient Research Centre. See the Case Study entitled "Dangerous Challenges and Hoaxes".

Prioritisation:

- TikTok considers Online Protection of Minors (and specifically the Mis-stated Age, Content, Contact and Conduct Risks) to be a **Tier 1** priority.
- Although TikTok's position is that the Platform is not specifically aimed at, or predominantly used by minors, it can be used by Younger Users. TikTok therefore treats the mitigation of online and real world harms to minors with the highest priority.
- TikTok will continue to implement targeted measures to protect minors, and dedicate specialist subject matter expertise and resources.

Key further risk mitigation effectiveness improvements in line with Art. 35 of the DSA:

- **Detection developments (Art. 35(1)(f)):** TikTok will continue to further improve its proactive detection and/or reporting capabilities, so that it can further improve its ability to detect Underage Users and those Younger Users who have mis-stated their age at registration.
- **Product developments (Art. 35(1)(a)):**
 - TikTok will continue to develop the use of private account by default for Younger Users aged 16-17;
 - TikTok will further develop its Underage User appeals functionality so that it can employ age estimation technology as an alternative to existing options for users to provide proof of age;
 - TikTok's content classification experts will develop plans to further expand Content Levels to additional content types and with new functionalities for users (see the Deep Dive entitled "Content Levels for Younger Users"); and
 - TikTok will further iterate and expand existing safety features, such as [screen time management](#) and [Family Pairing](#), as well continuing its focus on media literacy to generate awareness of existing tools and fostering the skills of users on how to create and engage with content on the Platform in a responsible and safe manner.
- **Continued risk monitoring and vigilance (Art. 35(1)(f)):** TikTok will continue to devote cross-functional resources on a priority basis to monitor and further mitigate these risks.

Risk Mitigations - Table 2	Online protection of minors and associated risks
TikTok's risk-mitigation measures in accordance with Art. 35(1) of the DSA (a) to (k), plus any other relevant measures:	
(a) Adaptation of feature or platform design, including online interfaces (as defined in Art 3(m) DSA):	<p>TikTok has made a number of adaptations to its platform to mitigate the risk relevant to minor safety as follows:</p> <p>To address the risk of Underage Users gaining access to the Platform, TikTok's current new user journey contains the following steps:</p> <ul style="list-style-type: none"> • In order to use the TikTok app, the app must first be downloaded from an app store. The TikTok app is rated 12+ in the Apple App Store and "Parental Guidance Recommended" in the Google Play Store, which empowers parents and guardians to prevent those under the minimum age requirement from downloading the app should they wish to do so using available parental controls on their device; • TikTok operates a neutral age gate, where the registration page does not include a statement that the platform is only for users aged 13 and over (which might be a prompt to certain users to guide their behaviour). If a prospective user enters a birth date indicating they are under 13, they will receive a non-eligibility notification [REDACTED]; • TikTok uses automated measures as a method to detect suspected Underage Users on the Platform and flag those accounts for review by [REDACTED]

trained moderators. [REDACTED]

[REDACTED] It employs automated review and also reporting mechanisms to detect suspected Underage Users and Younger Users using livestreams. Once detected, such accounts are sent to a specialist underage moderation process for human review;

- TikTok offers any users and non-users the ability to report other users that they suspect may be under 13 using a specific reporting reason entitled “*User could be under 13 years old*”, which again triggers the human moderation process;
- TikTok’s moderation systems and team will review reports received through the means set out above to identify users who are potentially aged under 13 years old. The relevant teams are trained and supervised to ensure a consistent, coherent and evidence-based approach;
- TikTok will immediately remove an account when it determines that the account holder is likely to be under the age of 13. The user is notified that this has happened and that such removal can only be reversed in the event of a successful appeal; and
- TikTok has structured the appeals process in consultation with users and external experts. TikTok offers a number of options for users to prove their age and TikTok’s process has also been designed to minimise data collection as far as possible.

To address the risk of Younger Users experiencing content that is not age-appropriate:

- TikTok employs ‘Content Levels’, a content classification framework that organises content based on thematic maturity and makes content with overtly mature themes unavailable to Younger Users. Content Levels operate to age-restrict content that does not violate the Community Guidelines, but which may not be age-appropriate for Younger Users. Under the [Youth Safety and Well-Being policy](#), TikTok’s Community Guidelines explain the role played by Content Levels, and set out the categories of content that are age-restricted to users aged 18 years and older;
- TikTok also applies warning labels and mask layers to certain videos, to provide users with additional information about the content so that they may decide in advance whether or not they wish to view it;
- TikTok employs a number of measures, including blocking keyword searches, to protect users (including Younger Users) from harmful material; [REDACTED]

[REDACTED] If these terms are searched for, TikTok redirects users to support materials;

- TikTok offers Family Pairing, which enables the account of a Younger User to be paired with that of a parent or guardian, so that the parent or guardian can manage certain settings relating to the content that the Younger User sees on the Platform, including to switch on Restricted Mode, switch off their teen’s ability to access the search functionality, and to filter out videos with words or hashtags from the FYF; and

	<ul style="list-style-type: none"> • TikTok offers all users a “Daily Screen Time” Management dashboard which includes a setting that allows users to manage their app usage. It lets users set a daily screen time limit reminder so they are notified when they reach that time on TikTok. For all Younger Users, the setting is turned on by default to 60 minutes. TikTok has incorporated the screen time dashboard and other screen time tools within Family Pairing. <p>To address the risk of Younger Users experiencing harm arising out of contact with other users:</p> <ul style="list-style-type: none"> • TikTok offers all users, including Younger Users the ability to report and block users who have engaged with their content in a way that is not appropriate or have posted harmful content; • For users aged 13-15: (1) accounts are defaulted to private at registration, meaning only people they have expressly approved as followers will be able to see their content, view their profile, or follow them; (2) other users can’t use a video from such a user in their own Duet or Stitch; (3) accounts are defaulted to allow only Friends to comment on users’ content and those users cannot allow Everyone to comment; and (4) TikTok does not display search results for such accounts unless the person conducting the search has [REDACTED]. • TikTok will continue to develop the use of private accounts by default for Younger Users aged 16-17. For public accounts (only) Everyone can comment, but users can change this setting if they wish. This does not apply to private accounts (where only people a user has expressly approved as followers will be able to see and comment on their content). In addition, all users can turn off ‘Allow comments’ to stop others commenting on their videos at any time. • For all Younger Users aged 13 to 17, TikTok: (1) does not suggest accounts belonging to under 18s to people aged 18+ (and vice versa); (2) offers tools to control who can comment on their content, filter and delete comments (including their own comments and comments posted by others on their videos); and (3) offers tools to make certain choices about who can watch and interact with videos they that they create. • Users must be aged 18 or older to host a livestream. TikTok employs a range of mitigation measures to moderate content from the host and those viewing. In addition, users must be at least 18 years old to send gifts to a creator during a livestream session.
<p>(b) Adaption of terms and conditions(as defined in Art 3(u) DSA) and their enforcement:</p>	<p>TikTok clearly sets a minimum age requirement, whereby users must be 13 years or older to have an account and to use the Platform, and this is reflected across a range of user-facing documentation and other resources. The Terms of Service say ‘<i>You can only use the Platform if you are 13 years of age or older. We monitor for underage use and we will terminate your account if we reasonably suspect that you are underage or are allowing someone underage to use your account. You can appeal our decision to terminate your account if you think we have made a mistake about your age. In short: You need to be 13 or over to use our Platform</i>’.</p>

	<p>TikTok's Terms of Service and Community Guidelines set out how TikTok works, what users can and can't do and the consequences of their use of the Platform. They state that <i>'We do not allow content that may put young people at risk of exploitation, psychological, physical, or developmental harm.'</i></p> <p>The Community Guidelines also detail the the types of content that are not permitted on the Platform at all and the content that is only available to users who have told us that they are over 18 years old, which is termed 'age-restricted content' which is described as content that includes: depiction of cosmetic surgery, activities that are likely to be imitated and may lead to any physical harm, significant body exposure of adults, seductive performances by adults, sexualised posing by adults, allusions to sexual activity by adults, blood of humans and animals, consumption of excessive amounts of alcohol by adults and consumption of tobacco products by adults.</p>
(c) Adoption of content moderation (as defined in Art 3(t) DSA) processes:	Please see the 'Key Information about TikTok' section for a description of TikTok's content moderation processes.
(d) Testing and adaption of algorithmic systems, including recommender systems (as defined in Art 3(s) DSA):	<p>Adaptations applicable to all types of content: TikTok's personalised FYF is one of the primary means through which users consume video content on the Platform. In addition to TikTok's content moderation policies and processes (see 'Key information about TikTok' above), further adaptations to mitigate the risk of violative content being recommended in the FYF are:</p> <ul style="list-style-type: none"> • TikTok detects and removes certain violative content at the point of attempted upload to the Platform or otherwise removes such content when detected. Such detected content cannot therefore be displayed on the FYF; • TikTok manually reviews video content when it reaches certain levels of popularity in terms of the number of video views, reducing the risk of violative content being shown in the FYF or otherwise being widely disseminated; and • TikTok offers users the tools to diversify the content displayed, to understand why videos have been recommended, to choose that certain keywords will not be displayed to them and to reset their FYF as if they were new to TikTok. <p>Relevant additional adaptations for minors are:</p> <ul style="list-style-type: none"> • In accordance with its approach to content moderation processes (described above), TikTok may take measures including to ensure that content will not appear in a user's FYF and may only be located through proactive search (which is subject to the mitigations described above). TikTok proactively detects harmful content using automated and human means, reducing the risk of age-inappropriate content being recommended or surfaced in a user's FYF. TikTok maintains content eligibility standards for the FYF that prioritise safety. The Community Guidelines set out categories of content that are not FYF eligible, which includes dangerous activities and challenges (see the Case Study entitled "Online Challenges and Hoaxes" below) TikTok considers that these measures also address

	<p>the risk of harmful content being widely disseminated on the platform via the FYF;</p> <ul style="list-style-type: none"> • An inherent challenge of any recommendation system is ensuring the breadth of content surfaced to a viewer is diverse and not too narrow or repetitive. TikTok employs mitigation measures to diversify content so that Younger Users are not exposed to repetitive content, which is especially important if they are exploring content related to more complex themes but which is not in violation of TikTok's terms or Community Guidelines; • TikTok offers a transparency tool to users called 'why this video' which explains why a particular video has been recommended to them in their FYF. Users can also refresh their personalised FYF on TikTok in order to view content on their FYF as if they just signed up for TikTok. Users can also use tools to automatically filter out videos with specific words or hashtags they do not want to see from their FYF or Following feeds and mark videos as "Not interested" so that similar videos will be shown less in the FYF; • Content posted by users under 16 is not eligible for FYF recommendation. Content from users over 16 is eligible for FYF recommendation. However, users can make their account or video private in order to prevent this; and • TikTok considers that the measures in relation to the design of the platform (see above in (a)) are also relevant, as they can be considered to reduce risks in connection with recommendation systems because they either restrict by default the dissemination of Younger Users' content and/or enable Younger Users to control the dissemination of their content on the Platform.
<p>(e) Adaptation of advertising systems:</p>	<p>Having regard to Art. 28(2) DSA, TikTok does not serve ads to Younger Users in Europe based on profiling. Instead, ads are based only on generic data points; for example, country and language.</p> <p>TikTok imposes a number of restrictions to ensure that the generic ad content seen by Younger Users is age appropriate. TikTok also does not display generic ads to Younger Users for products and services which are not suitable for them, such as energy drinks and video games or other media content with an age rating of more than 13.</p>
<p>(f) Reinforcing risk detection measures:</p>	<p>TikTok operates a number of specialist teams that deal with illegal content and minor safety content in particular and who play a role in reinforcing risk detection measures:</p> <ul style="list-style-type: none"> • TikTok has a specialist Youth Safety and Wellbeing Issue Policy Team comprised of 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 overarching goal of the team referred to above is to best protect young people's unique developmental life stages while accounting for diverse global experiences;

	<ul style="list-style-type: none"> • The team works closely with the Product Policy Teams and Trust & Safety Product teams on youth-specific strategies to address harms captured across all Trust & Safety policies and feature designs; and • Within the Trust & Safety Product Team, there is a dedicated function for livestream safety.
(g) Cooperation with trusted flaggers:	Trusted Flaggers are not yet relevant in the context of this content. However, TikTok operates a Community Partner Channel whereby onboarded NGOs perform a similar role to designated trusted flaggers (under the DSA) to submit reports of suspected harmful content. These NGOs can then report suspected harmful content directly to TikTok's Trust & Safety teams, who review such reports on a priority basis.
(h) Cooperation with other platforms through the codes of conduct/crisis protocols:	Not applicable.
(i) Awareness-raising measures for recipients of the services:	<p>TikTok has undertaken the following example programmes:</p> <ul style="list-style-type: none"> • Saferinternet.at (Austria): TikTok has partnered with Saferinternet.at for an in app activation or Safer Internet Day on 7 February 2023, focusing on TikTok's Community Guidelines, safety features and privacy settings and promoting their hashtag #SID2023AT; • Fondazione Carolina (Italy): TikTok has been partnering with Fondazione Carolina, the NGO named after the first victim of cyberbullying in Italy, on the project "Parents in Blue Jeans". This project involves a series of local events to raise awareness, both among students and parents, on digital wellbeing, cyberbullying, conscious use of TikTok, and parental control through a dedicated Guide for Parents; and • MIND (Sweden): TikTok has worked closely with MIND Together to create a campaign to empower users around digital wellbeing (#5stegframåt).
(j) Targeted measures to protect the rights of the child:	<p>TikTok considers it necessary to strike a balance between safety and the rights and freedoms of the child. In particular, by way of non-exhaustive example, TikTok has reference to:</p> <ul style="list-style-type: none"> • The Charter and in particular Art. 24 on the rights of the child which states that <i>"In all actions relating to children, whether taken by public authorities or private institutions, the child's best interests must be a primary consideration"</i>; • Art. 13 of the UN Convention on the Rights of the Child enshrines the right of children to freedom of expression, including <i>"freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of the child's choice."</i> ; and • Art. 17 of the Convention which protects the rights of children to <i>"access appropriate information"</i>, and specifically encourages: <i>"... the mass media to disseminate information and material of social and cultural benefit to the child and in accordance with the spirit of article 29 [of the Convention]; "... international co-operation in the production, exchange and dissemination of such information and</i>

	<p><i>material from a diversity of cultural, national and international sources; and “... the production and dissemination of children’s books”.</i></p> <p>TikTok acknowledges that a balance must be found between the design of measures to address the Mis-Stated Age Risk and the fundamental rights of legitimate users of any age to receive information and to freedom of expression, including the right to express themselves freely on the Platform. Accordingly, TikTok has to strike the right balance and avoid unreasonable and disproportionate impacts on such legitimate users (i.e., potentially resulting in preventing them from accessing the Platform, such as by creating unreasonable burdens during sign up or excessive removal of suspected underage accounts in error, and related privacy impacts, that could function as a deterrent to legitimate users).</p>
<p>(k) Measures to identify and address inauthentic content and behaviours:</p>	<p>There is a risk that bad actors may intentionally manipulate the Platform or make inauthentic use of the Platform. TikTok takes a range of measures to prevent and mitigate the risk, including:</p> <ul style="list-style-type: none"> • Account impersonation may enable users to craft identities that appear trustworthy and relatable to their intended audience. Users are not allowed to use someone else’s name, biographical details, or profile picture in a misleading manner. For example, impersonation may enable hateful organisations to craft identities that appear trustworthy and relatable to their intended audience. TikTok uses a range of methods to detect and remove these accounts; and • TikTok does not allow coordinated attempts to influence or sway public opinion while also misleading individuals, the community, or its systems about an account’s identity, approximate location, relationships, popularity, or purpose. TikTok investigates and removes these operations, focusing on [REDACTED] to determine if actors are engaging in a coordinated effort to mislead TikTok’s systems or its community.

Case Study: Online challenges and hoaxes

TikTok's research conducted with Praesidio Safeguarding to understand risks and design detailed solutions

Context: This case study explains how TikTok worked with academic experts and conducted international research with minors to better understand risks relating to young people's engagement with potentially harmful online challenges and hoaxes, and took an evidence-based approach to assessing detailed solutions.

An evidence-based approach to online challenges and hoaxes: To better understand young people's engagement with potentially harmful challenges and hoaxes, TikTok commissioned a global research project in 2021, in which TikTok:

- surveyed more than 10,000 teenagers, parents and educators across 10 countries;
- convened a panel of 12 leading youth safety experts; and
- partnered with a leading clinical child psychiatrist and a behavioural scientist specialising in risk prevention.

TikTok commissioned Praesidio Safeguarding, an independent safeguarding agency, to write a report to capture key findings and recommendations. The report, entitled "[Exploring effective prevention education responses to dangerous online challenges](#)" was written by Dr. Zoe Hilton, Director and Founder of Praesidio Safeguarding (and was made publicly available through a [newsroom post](#)).

Key findings: The research project involved listening to teenagers, parents and educators, and revealed the following key themes:

Challenges:

- Most teenagers regard most challenges as fun or safe and only 0.3% of teenagers had taken part in an online challenge they categorised as very dangerous;
- Teenagers want both more and better information to assess risk presented by online challenges; and
- Teenagers have sought support and advice about challenges, but some parents and teachers find it difficult to discuss.

Hoaxes:

- People don't know how to assess hoaxes and only a minority are able to identify them as clearly fake;
- Hoaxes are experienced more negatively than challenges. This can impact on users' mental health; and
- Parents and educators are concerned but feel ill equipped to support teenagers with hoaxes.

Measures taken by TikTok in response to the report's findings: The findings from the report informed a review of TikTok's policies and processes that protect against both dangerous acts and suicide or self-harm hoaxes to evaluate what it could do to strengthen its existing protections. As part of this review TikTok implemented the following measures:

- **Community Guidelines:** TikTok updated its Community Guidelines (to create a new standalone policy on [Dangerous Activities and Challenges](#)), and made changes to its moderation practices.

- **Safety Centre:** One of the main findings from the report is that teenagers, parents, and educators need better information about challenges and hoaxes. As a result, TikTok has worked with experts including The Net Safety Collaborative to develop a new resource for its Safety Centre dedicated to [Online Challenges](#), which includes advice for parents/caregivers to help address the uncertainty they expressed about discussing this topic with their teenagers. The resource is available in a range of European and other languages, including Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Italian, Latvian, Lithuanian, Norwegian, Polish, Portuguese, Romanian, Russian, Slovak, Spanish, Swedish, and Ukrainian.
- **Safety Messaging:** TikTok worked with 38 creators in 13 markets to develop safety videos, which call on the community to follow the 4-step process (Stop, Think, Decide, Act) when engaging with an online challenge. These videos were launched in February 2022 and spotlighted in a dedicated [#SaferTogether hub](#) on the Discovery page. TikTok also deployed an algorithm to ensure that users on the Platform under the age of 18 will begin to see these videos in their FYF.
- **Enforcement:** In response to findings in the report that sharing alarming warnings about hoaxes can cause panic and harm, TikTok adapted its approach to enforcement of its Community Guidelines by removing content that includes warnings about suicides or self-harm hoaxes as they misleadingly imply the hoax is real.
- **Content labels:** TikTok worked with experts to improve the language used in its warning labels that would appear to users who attempt to search for content related to harmful challenges or hoaxes.
- **Prompt:** TikTok introduced a new prompt to encourage community members to visit its Safety Centre to learn more, and should people search for hoaxes linked to suicide or self-harm, TikTok began displaying additional help and educational resources in search.

Deep Dive: Content levels for Younger Users

An explanation by TikTok of its design and delivery system to categorise content in order to protect minors

Overview: Content Levels is a key part of TikTok's commitment to ensure that Younger Users have an age-appropriate experience on the Platform. Content Levels is a content classification framework that organises content based on thematic maturity and makes content with overtly mature themes unavailable to Younger Users. Content Levels operate to age-restrict content that does not violate the Community Guidelines such that it would be banned, but which may not be age-appropriate for Younger Users. Under the [Youth Safety and Well-Being policy](#), TikTok's Community Guidelines explain the role played by Content Levels, and set out the categories of content that are age-restricted to users aged 18 years and older.

Building the framework: The development of Content Levels has been led by the Content Classification policy team, who are experts in content classification and maturity ratings. The team has deep industry 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 digital entertainment; and developmental psychology. The principles underpinning Content Levels have been developed based upon the long established approach taken by content classification and ratings bodies in the film industry, television and gaming industries, with certain adaptations for TikTok.

How it works: As the [Overview](#) of TikTok's Community Guidelines explains, TikTok removes violative content from the Platform. Where content is not violative, *but TikTok detects that it contains mature or complex themes (e.g. fictional scenes that may be too intense for Younger Users), a maturity rating is applied to it. This indicates that the content is only suitable for audiences aged 18 or older, and ensures that Younger Users are protected from such content.*

Categories of age-restricted content: The [Youth Safety and Well-Being policy](#) of the Community Guidelines lists the categories of content currently age-restricted to users aged 18 and older:

- *Cosmetic surgery that does not include risk warnings, including before-and-after images, videos of surgical procedures, and messages discussing elective cosmetic surgery;*
- *Activities that are likely to be imitated and may lead to any physical harm;*
- *Significant body exposure of adults;*
- *Seductive performances by adults;*
- *Sexualised posing by adults;*
- *Allusions to sexual activity by adults;*
- *Blood of humans and animals;*
- *Consumption of excessive amounts of alcohol by adults;*
- *Consumption of tobacco products by adults.*

7. RISKS TO ELECTIONS AND CIVIC INTEGRITY

Description of the risk:

- 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**”).
- 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:
 - Misleading information about how to vote or register to vote or the qualifications to vote or run in an election;
 - 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);
 - Misleading information about how to participate in a census or eligibility requirements for participating in a census;
 - Content that advances false claims related to the technical eligibility requirements for current political candidates and sitting elected government officials to serve in office;
 - False claims of election fraud (such as voting machines being tampered with to favour a candidate or political party);
 - Content that falsely claims that an election has been or will be rigged, so the results cannot be trusted;
 - Misinformation or/and conspiracy theories about candidates, candidate impersonation and related issues that may impact on civic integrity; or
 - 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 on civic integrity if mis-used for political endorsements or other purposes (TikTok refers to this as “**Synthetic and Manipulated Media**”).
- TikTok also acknowledges that Arts. 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.

Risk mitigation measures put in place:

- Risk Mitigations - Table 3 sets out a summary of risk mitigation measures in place with reference to Art. 35 (1) (a) to (k) of the DSA.
- Key mitigations for this risk are: (1) TikTok’s cross functional Election Integrity Programme which covers both national member state and EU parliamentary elections; (2) a Fact-Checking Programme, with [IFCN-certified](#) independent fact-checking partners across Europe who assist TikTok in verifying, labelling or removing content; (3) in-app interventions for users to direct them to authoritative content; and (4) proactive detection and interventions in relation manipulated content, covert influence operations and prohibited political ads.
- TikTok gives appropriate consideration to balancing freedom of expression and user’s rights to participate in political discourse and process when designing mitigation measures. TikTok also strives to balance addressing misinformation with users’ rights to express their personal opinions about politicians and electoral processes.

- As risks relating to Election Misinformation tend to be highly localised, TikTok's Election Integrity Programme makes sure that these differences are understood and factored into planning and execution of risk mitigation strategies.

Key data relied on:

- TikTok does not separately report on content detected and removed from the Platform due to Election Misinformation and content removed for these reasons is contained within its reporting on harmful misinformation/content.
- TikTok reported in its Q1 2023 Community Guidelines Enforcement Report that it detected and removed 908,927 pieces of video content globally under TikTok's policies and processes for integrity and authenticity (which comprises harmful misinformation/content as well as spam and fake engagement). In addition:
 - 94.8% of such content was detected and removed proactively, before any reporting;
 - 72.8% was removed before there were any views of that content; and
 - 76.6% was removed within 24 hours of upload.
- The same report shows that during Q1 2023, TikTok's teams detected and removed a number of covert influence operations that sought to artificially amplify certain viewpoints in the context of election cycles, however none of these related to elections in Europe.
- TikTok's reporting under the Code of Practice on Disinformation (the "CoPD") contains relevant information and metrics on its efforts to combat disinformation:
 - In Q1 & Q2 2023, fewer than 2 in 10,000 views occurred on content identified and removed for violating its policies on harmful misinformation (all, not just Election Misinformation): i.e., across the EU, 140,635 videos were removed, with the number of views of such videos in the EU being 1,012,020,899 (the equivalent numbers for the EEA are 142,711 and 1,019,752,855 respectively).¹⁸
 - The number of ads removed for violating TikTok's ads policies under the political content ad policy was 390 in the EU/395 in the EEA (for the 6-month period of 1 January to 30 June 2023, as indicated in its CoPD Report).¹⁹

Severity:

- The spread of Election Misinformation in the context of a European election could potentially have a societal impact in a relevant member state, region or at the EU level, such as by influencing the environment surrounding an election or other civic event, which could have a undermine or erode trust in election processes and institutions.
- The scale of Election Misinformation risks tend to be localised by reference to where the election is taking place. Risks relating to Election Misinformation are likely to arise in that location for a largely predictable period of time around an election, which are generally scheduled several months in advance, which facilitates proactive scenario-planning and risk management.
- Such risks can generally be remediated by implementing effective measures to proactively detect and remove election misinformation, through fact-checking and by labelling such content, as well as reducing its negative impact through in-app interventions to lead users to reliable information (such as dedicated election information hubs).
- TikTok's key data above demonstrates that TikTok's mitigation measures have a significant impact on the scale and potential duration of harm to users from harmful misinformation/content generally.
- TikTok assesses the potential severity of risk of Election Misinformation to be moderate due

¹⁸ For the period 1 January - 30 June 2023, as reflected in TikTok's Code of Practice on Disinformation, p. 135 - 137.

¹⁹ For the period 1 January - 30 June 2023, as reflected in TikTok's Code of Practice on Disinformation, p.4 - 6 (to be

to: (1) the nature of the harm, whilst acknowledging that the general severity of the effects of misinformation are impacted by a range of factors and variables, such as the characteristics of the individual, susceptibility to misinformation and access to other information sources, and wider societal factors; and (2) TikTok's proactive efforts to enforce its policies, as demonstrated by the mitigation measures referred to above and in particular to TikTok's 94.8% proactive detection rate. In particular this assessment results from the measures put in place to restrict the scale of the dissemination of harmful misinformation/content, the duration for which it may circulate on the Platform, and the measures put in place to prevent repeated viewing of such content.

Probability:

- TikTok's Election Integrity Programme tracks elections up to 15 months in advance. TikTok is therefore aware of the forthcoming elections in the EU, which include elections in Slovakia, Poland, the Netherlands and the 2024 EU Parliamentary elections. TikTok sets out an overview of measures taken in the context of the recent Spanish election in the Case Study below.
- TikTok strives to prevent the upload of, or otherwise remove, Election Misinformation from the Platform. TikTok assesses that it is possible that there will remain some level of Election Misinformation on the Platform, which is likely to occur at specific times (i.e. during election cycles). This is due to the cyclical yet predictable nature of Election Misinformation, making it possible to anticipate and plan for effective risk management. This conclusion is supported by the volumes reported above as well as TikTok's Election Integrity Programme tracking.
- TikTok does not permit political ads on its Platform and does not permit government politicians and political party accounts to monetise political content. TikTok has a number of Ad Policies in place which are specifically designed to protect users from misleading information in the context of elections. TikTok's pre-moderation activities means that any ads on the Platform are unlikely to include Election Misinformation.

Key stakeholder engagement:

- As part of its Elections Integrity Programme, TikTok seeks to iterate and improve its policies and processes by engaging with civil society and national election authorities to understand local contexts and to obtain authoritative information regarding the conduct of elections under their remit.
- As part of its work on the CoPD and to build on its work to combat disinformation (including Election Misinformation) representatives will actively co-chair and participate in the CoPD Elections Working Group.
- As part of TikTok's Fact-Checking Programme, TikTok works closely with its [IFCN-certified](#) independent fact-checking partners (TikTok's European fact-checkers include Agence France-Presse (AFP), Deutsche Presse-Agentur (DPA), Facta, Logically, Lead Stories, Newtral, Science Feedback, Teyit, and Reuters).

Prioritisation:

- TikTok considers the risk of dissemination of Election Misinformation content to be a **Tier 1** priority.
- This is due to the online and potential real world societal harms and civic impacts of the significant schedule of upcoming EU elections, the resulting risk from which could be dynamic in nature. TikTok therefore treats Election Misinformation as a key priority for its risk mitigation endeavours.

- TikTok will implement targeted measures and dedicate specialist resources (including working with TikTok's [IFCN-certified](#) fact-checking partners across Europe) in advance of European elections as part of its Elections Integrity Programme, including forthcoming elections in Slovakia, Poland, the Netherlands and the 2024 EU Parliamentary election.

Key further mitigation effectiveness improvements in line with Art. 35 of the DSA:

- **Detection improvements (Art. 35(1)(f)):** TikTok will further expand its Fact-Checking Programme by on-boarding new European-based fact-checking partners and increasing its operational coverage in the EU.
- **Continued risk monitoring and vigilance (Art. 35(1)(f)):** TikTok will continue to devote cross-functional resources on a priority basis to its processes for handling Election Misinformation risks.

Risk Mitigations - Table 3	Risks to elections and civic integrity
TikTok's risk-mitigation measures in accordance with Art. 35(1) of the DSA (a) to (k), plus any other relevant measures:	
(a) Adaptation of feature or platform design, including online interfaces (as defined in Art. 3(m) DSA):	<p>TikTok implements a range of measures to prevent and mitigate the risk of the dissemination of violative content on or through the Platform, including through the following features, Platform design, and relevant technical safety measures:</p> <ul style="list-style-type: none"> • TikTok 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. The content is ineligible for the FYF and search restrictions are also applied; • TikTok will prompt users to (re)consider unverified content before sharing it to other users on the Platform; • TikTok may deploy a custom video notice tag to alert all users encountering videos containing premature election results claims that the result had not yet been officially declared; • TikTok may, from time to time, block certain search results associated with Election Misinformation from appearing on the Platform; • TikTok labels as 'state controlled media' those accounts that it determines are run by media entities whose editorial output or decision-making process is subject to control or influence by a government. This label is available in 22 official EU, and other global, languages; and • TikTok has a range of measures in place to mitigate the risk of a livestream being capable of containing Election Misinformation.
(b) Adaption of terms and conditions(as defined in Art. 3(u)	<p>TikTok prohibits Election Misinformation on the Platform through a combination of TikTok's Terms of Service and the Community Guidelines, and takes a range of measures to generate awareness and inform users.</p> <p>Under the heading of "Integrity and Authenticity", TikTok's Community</p>

<p>DSA) and their enforcement:</p>	<p>Guidelines prohibit Election Misinformation, stating that ‘We do not allow misinformation about civic and electoral processes, regardless of intent’. TikTok also states that ‘We do not allow paid political promotion, political advertising, or fundraising by politicians and political parties (for themselves or others)’. The Community Guidelines also make it clear that Government, Politician, and News Accounts will be treated like any other account in the context of moderation activities.</p> <p>The Guidelines provide further detail to users about what each of these terms mean and make clear that statements of personal opinion (as long as it does not include harmful misinformation/content) are permitted.</p>
<p>(c) Adoption of content moderation (as defined in Art. 3(t) DSA) processes:</p>	<p>Please see the ‘Key Information about TikTok’ section for a description of TikTok’s content moderation processes.</p>
<p>(d) Testing and adaption of algorithmic systems, including recommender systems (as defined in Art. 3(s) DSA):</p>	<p>TikTok’s personalised FYF is one of the primary means through which users consume video content on the Platform.</p> <p>In addition to TikTok’s content moderation policies and processes (see ‘Key information about TikTok’ above), further adaptations to mitigate the risk of violative content being recommended in the FYF are:</p> <ul style="list-style-type: none"> • TikTok detects content that may be Election Misinformation and if it cannot be determined by a fact checker whether it is/not such content, TikTok may label it as ‘unverified content’. The content will then be ineligible for the FYF; • The Elections Integrity Programme team will conduct sampling of election-related content in the run-up to elections to detect risks, which may involve conducting an assessment of the top reported videos associated with the election; • TikTok manually reviews video content that reaches certain levels of popularity in terms of the number of video views, reducing the risk of violative content being shown in the FYF or otherwise being widely disseminated; and • TikTok offers users the tools to diversify the content displayed, to understand why videos have been recommended, to choose that certain keywords will not be displayed to them and to reset their FYF as if they were new to TikTok.
<p>(e) Adaptation of advertising systems:</p>	<p>Please see the ‘Key Information about TikTok’ section for a description of TikTok’s content moderation processes.</p> <p>In addition with specific applicability to Election Misinformation:</p> <ul style="list-style-type: none"> • Restrictions for Government, Politician, and Political Party Accounts (“GPPAs”): TikTok’s priority in this space is to keep harmful election misinformation off the Platform and ensure its community has a positive and welcoming experience. TikTok’s political ads policy explains how and why holders of GPPAs are prevented from accessing monetisation features, including TikTok’s paid ads services. Solicitations for campaign fundraising by GPPAs are also not permitted on the Platform;

	<ul style="list-style-type: none"> ● Prohibition on political ads: TikTok has long prohibited political ads, including both paid ads and creators being paid to make branded political content. For example, ads that reference an election, including voter registration, voter turnout, and appeals for votes and that promote or attack government policies or track records are prohibited. Accounts that TikTok identifies as belonging to politicians and political parties have their access to advertising features turned off. This includes candidates or nominees for public office, the spouses of candidates and Royal Family members with official government capacities. All ads are reviewed before being displayed on the Platform. Upon review, if an ad is deemed to violate TikTok's prohibition of political ads, it will not be permitted on the Platform. TikTok also reviews ads reported to it and upon review, any violating ads will be removed; and ● Campaign (Election) Fundraising: TikTok does not permit solicitations for campaign fundraising by GPPAs on its 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. During election periods, TikTok increases its risk mitigation measures for political content due to the risk of increased political activity on the Platform.
<p>(f) Reinforcing risk detection measures:</p>	<p>TikTok operates a number of specialist teams that deal with Election Misinformation in particular and who play a role in reinforcing risk detection measures:</p> <ul style="list-style-type: none"> ● TikTok's Election Integrity Programme entails a highly cross-functional approach, with subject matter experts across both its Trust & Safety team and multiple other teams. Specifically this involves its Trust & Safety Integrity & Authenticity Policy, Product, OPM, Incident Management, Risk Analysis, and Operations teams. Outside of Trust & Safety, the Programme involves TikTok's global and regional Legal teams, its regional Public Policy and Government Relations teams, Communications teams, and others. There is also a close working relationship with TikTok's Monetisation Integrity team. The team design and operate a programme for each election and maintain a real time monitoring dashboard of evolving information; ● TikTok's Fact-checking Programme is led by members of the Trust & Safety teams referenced below 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 the Integrity & Authenticity Policy, Product and OPM teams), and multiple other teams. The Fact-checking Programme is a core part of TikTok's strategy to tackle harmful misinformation on a cross-team and cross-functional basis; and ● TikTok's Political and Regulatory Risk Team and Incident Response Team conduct market specific risk assessments up to 15 months prior to an election taking place. These teams work to ensure proper risk mitigation steps are met, including language resources, potential risks posed to business in each market, and compliance with all market specific political advertising regulations.

(g) Cooperation with trusted flaggers:	<p>TikTok operates a Community Partner Channel whereby onboarded NGOs perform a similar role to trusted flaggers to submit reports of suspected harmful content. In advance of an election, TikTok assesses whether there may be coverage gaps for purposes of flagging Election Misinformation. Based on this assessment its OPM team will engage with and seek to onboard suitable NGOs to the Community Partner Channel in advance of an election.</p> <p>TikTok is a signatory to the CoPD and is a participant in the Permanent Taskforce and subgroups set up under the Code. TikTok has produced two reports from the period of Q4 2022 and Q1 & Q2 2023. The Permanent Taskforce's approach has created a framework for platforms, NGOs and other ecosystem stakeholders to effectively work together.</p>
(h) Cooperation with other platforms through the codes of conduct/crisis protocols:	<p>TikTok is a signatory to the EU Code of Practice on Disinformation and is a participant in the Permanent Taskforce and subgroups set up under the Code. TikTok has produced two reports from the period of Q4 2022 and Q1 & Q2 2023 including how it combats Election Misinformation and promotes election and civic integrity. The Permanent Taskforce's approach has created a framework for platforms, NGOs and other ecosystem stakeholders to effectively work together (including a newly formed Working Group on Elections, of which TikTok is a co-chair).</p>
(i) Awareness-raising measures for recipients of the services:	<p>TikTok undertakes the following measures: (1) TikTok's Safety Centre includes a page on Election Integrity, which explains its approach to ensuring election integrity; (2) TikTok's Transparency Centre webpage, in the 'Our Commitments' page, contains articles explaining its approach to areas of Election Misinformation; (3) TikTok may create a user facing elections hub within its app ahead of certain elections taking place, which would then appear in search results related to the election or can be linked to with a hashtag; (4) TikTok may conduct on and off app media literacy campaigns. Please see the Case Study entitled "Spain's parliamentary election (July 2023)" below.</p>
(j) Targeted measures to protect the rights of the child:	<p>Not applicable.</p>
(k) Measures to identify and address inauthentic content and behaviours:	<p>Please see the corresponding section of the Hate Speech section of this Report.</p> <p>In addition with specific applicability to Election Misinformation:</p> <ul style="list-style-type: none"> • TikTok does not allow the use of accounts to engage in Platform manipulation. This includes the use of automation to register or operate accounts in bulk, to distribute high-volume commercial content, to artificially increase engagement signals, or to circumvent enforcement of TikTok's policies; • Synthetic media or manipulated content that shows realistic scenes must be disclosed and TikTok does not allow synthetic or manipulated content that contains the likeness of any real private figure. TikTok does not allow material that has been edited, spliced, or combined (such as video and audio) in a way that may mislead a

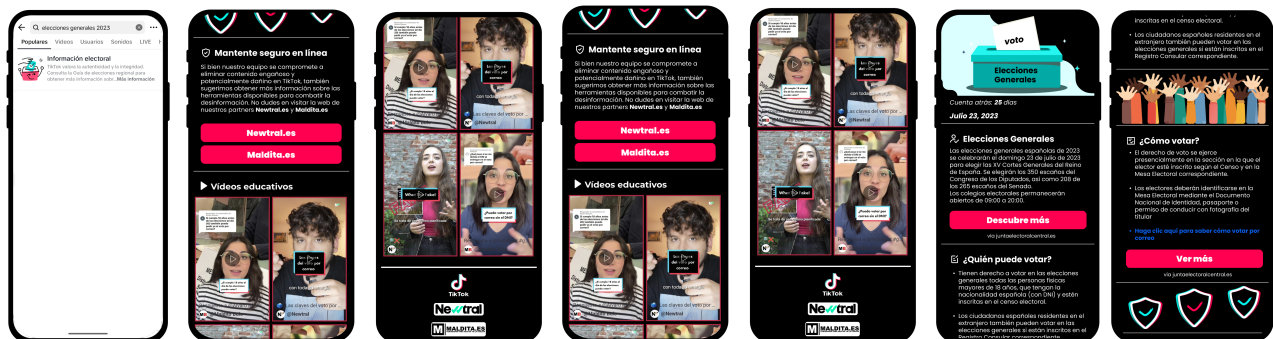
person about real-world events. TikTok does allow synthetic media of public figures as long as the content is not used for an endorsement or is otherwise violative;

- TikTok implements a number of methods to monitor for inauthentic accounts in order to detect larger clusters or similar inauthentic behaviours. This includes monitoring cross platform threats, leveraging internal detection signals, and manually investigating reports from users and trusted flaggers. If any anomalous activity is detected, TikTok will investigate these accounts further to determine whether there is sufficient evidence of an inauthentic activity. These accounts and the behaviours associated with them (such as likes, follows) may automatically be removed from the Platform if certain thresholds are met;
- 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;
- TikTok offers political account verification to protect users from impostor accounts that may attempt to spread civic misinformation while pretending to represent a political party or a politician. TikTok applies a [verified badge](#) to such accounts, which is the blue check mark symbol that appears next to a user's account name in search results and on the profile;
- TikTok does not allow coordinated, inauthentic, or adversarial behaviours aimed at misleading its users, undermining the integrity of the public conversation, or manipulating its platform for financial, political, or ideological purposes. When TikTok investigates and removes these operations, it focuses on [REDACTED] to determine if actors are engaging in a coordinated effort to mislead TikTok's systems or its community. It uses several types of information (including open-source and proprietary) to assess covert influence operations and leverages a standard framework of confidence assessment to help ensure it makes consistent and accurate determinations; and
- TikTok's [state-affiliated media policy](#) is designed to label accounts run by entities whose editorial output or decision-making process is subject to control or influence by a government.

Case Study: Spain's parliamentary election (July 2023)

Context: Elections are unique events that may present novel policy issues and require decisive but thoughtful action from TikTok's teams under tight timelines. TikTok is committed to delivering 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 work, which was implemented in the recent Spanish election which took place on 23 July 2023 and is being implemented for the upcoming election in Slovakia. Specific measures taken during the Spanish 2023 election were:

- **In-app election hub:** In advance of the Spanish election, TikTok sought to address misinformation risk by providing users with access to an in-app election hub, which contains information about the election from reliable sources. The key objective of the in-app election hub is to inform users about the election process, including: (i) when to vote; (ii) where to vote; and (iii) the process of voting.
- **Media literacy:** Maldita, a local media literacy organisation, produced educational videos about the electoral process and election misinformation which were made available on the hub. Maldita is a verified signatory of the International Fact-Checking Network's ("IFCN") code of principles and member of the European Fact-Checking Standards Network ("EFCNSN").
- **Fact-checkers:** TikTok has fact-checking coverage in 17 official European languages, including Spanish. As part of the 2023 Spanish election, TikTok partnered with Newtral to help support its election integrity efforts in Spain and tackle any election misinformation on the platform. Throughout the election, Newtral proactively monitored content on the Platform to identify potential misinformation and provided reports on trends observed both on the Platform and off the Platform.
- **Speaker series:** TikTok ran an Election Speaker Series in advance of the 2023 Spanish election. As part of the Election Speaker Series, TikTok invites suitably qualified external local/regional experts to share their insights and market expertise with its internal teams in order to inform its approach to the upcoming election. As part of TikTok's preparations for the Spanish election, the local fact-checking partner, Newtral, was engaged to provide a Speaker Series presentation.
- **Media literacy:** In addition to providing access to reliable information through the in-app election hub, as part of the 2023 Spanish election, TikTok worked with Maldita to produce educational videos about the electoral process and election misinformation.



8. RISKS OF GENDER-BASED VIOLENCE CONTENT

Description of the risk:

- TikTok understands that neither the DSA nor EU law contains a single or comprehensive definition of gender-based violence (“**GBV**”). TikTok therefore understands GBV 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”*.^{20 21}
- The EU Commission has adopted a proposal for a directive on combating violence against women and domestic violence, which includes a proposed definition of “violence against women”.²² In July 2023, the European Parliament adopted its position on the directive.²³
- 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;
 - 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 Art. 21 of the Charter prohibits discrimination based on sex and sexual orientation, and that Art. 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.²⁴

Key mitigation measures put in place:

- Risk Mitigations - Table 4 sets out a summary of risk mitigation measures in place with reference to Art. 35 (1) (a) to (k) of the DSA.
- Key mitigations for this risk are: (1) TikTok’s policies and proactive content moderation enforcement systems and processes; (2) the limits placed on searching for violative content; (3) incident management and risk detection activities including liaison with specialist law enforcement units; and (4) specific actions taken in relation to the For You recommender system.

²⁰ [The Explanatory Report to the Council of Europe Convention on preventing and combating violence against women and domestic violence](#).

²¹ Note that we consider violative content related to hate speech against people due to their sexual orientation in the separate Hate Speech section of this Report.

²² European Commission, [Ending gender-based violence](#).

²³ European Parliament, [Combating violence against women: MEPs ready to negotiate on draft EU directive](#) (12 July 2023).

²⁴ [Proposal for a Directive on combating violence against women and domestic violence, Recital 8](#).

- An assessment of the relevant local/regional, cultural and linguistic factors are integrated as part of TikTok's Trust & Safety team's risk management processes, which includes ensuring that TikTok is up to date on the variations in gender-related hate speech and localised trends.

Key data relied on:

- As noted above, GBV may involve a wide range of content and behaviours. Taking harassment and bullying as examples of GBV, TikTok reported in its Community Guidelines Enforcement Report from Q1 2023 that it detected and removed 4,987,800 pieces of video content globally under TikTok's policy and processes for harassment and bullying, which includes content removals under TikTok's Community Guidelines which prohibit: (1) sexual harassment; (2) threats of hacking, doxxing and blackmail; and (3) abusive behaviour. In addition:
 - 85.3% of such content was detected and removed before it was reported to TikTok;
 - 58.9% of such content was detected and removed before it has received any views by users of the Platform; and
 - 79.5% of such content was removed within 24 hours of being posted on the Platform.

Severity:

- The negative effects of the dissemination of GBV Content can compound serious psychological and physical trauma with severe consequences for victims. Depending on the nature of the content, it may have substantial negative effects for the groups or individuals that it targets, as well as negative impacts on society more generally by undermining human rights principles of freedom of expression, tolerance and non-discrimination, potentially through indoctrination of hateful ideologies and discriminatory views (such as misogyny).
- For individuals or groups that are the targets or victims of GBV, they risk suffering prejudice, discrimination and violation of their human rights, including but not limited to the right to human dignity. At the more extreme end, the dissemination of GBV Content increases the risk of real-world violence against such groups. In terms of scale, such risks can have a societal impact in specific areas of, or in individual EU member states within, the EU.
- The impacts of GBV Content will depend on its nature and can range in remediability. However, these impacts may be mitigated through strong enforcement measures, product design and interventions, and awareness-raising measures.
- TikTok's key data above demonstrates that TikTok's mitigation measures have a significant impact on the scale and potential duration of harm to users from GBV (as included within reporting on video removals for bullying and harassment).
- TikTok assesses the potential severity of the risk of dissemination of GBV content to be material, due to: (1) the nature of the harm involving potentially serious psychological and physical trauma, real-world violence, and impact on freedom of expression, tolerance, and non-discrimination, at a societal level; and (2) TikTok's proactive efforts to enforce its policies, as demonstrated by the mitigation measures referred to above and in particular to TikTok's proactive detection rates for content including GBV, as shown above. In particular this assessment results from the measures put in place to restrict the scale of the dissemination of GBV, the duration for which it may circulate on the Platform, and the measures put in place to prevent repeated viewing of such content.

Probability:

- GBV may be manifestly illegal on its face or it may take other forms. TikTok updated its Community Guidelines in March 2023 to recognise GBV as a separate focus for its moderation and mitigation efforts. While TikTok already prohibited male supremacy ideology as a form of hate speech, TikTok made this decision to bring more transparency to its policies following the high profile incident in 2022 where GBV content was repeatedly uploaded to the Platform promoting the male supremacy ideologies of Andrew Tate. GBV will now be tackled under a new sub-policy called [Sexual Exploitation and Gender-Based Violence](#) in order to further assist TikTok's understanding of the risk.
- TikTok strives to prevent the upload of GBV content on the Platform. TikTok proactively enforces its policies on GBV and operates reporting channels for users. The key data above demonstrates that TikTok's mitigation measures have a significant impact on reducing the likelihood of dissemination of GBV content on the Platform.
- TikTok strives to prevent the upload of, or otherwise remove, GBV from the Platform. TikTok considers that it is possible that there will remain some level of dissemination of GBV content on the Platform. However, TikTok will gain a deeper understanding once it is able to analyse the results of its Q2 Community Guidelines transparency reporting, where GBV content moderation actions will be reported for the first time.
- TikTok's pre-moderation of ads means any advertising on the Platform is unlikely to include forms of GBV content.

Key stakeholder engagement:

- TikTok works with various organisations focused on combating gender-based violence, and integrates safety features and resources into the Platform. This includes [StopNCII.org](#), a global organisation focused on stopping non-consensual image abuse, which is a user safety resource in its [online Safety Center](#).
- TikTok also works with SOS Viol (Belgium), Danner (Denmark), an organisation working to eliminate violence against women through empowerment, protection, prevention, and advocacy, the Rape Crisis Centre Tukinainen (Finland), the Dublin Rape Crisis Centre (Ireland) and Kvinnofridslinjen (Sweden).
- TikTok also works closely with organisations that support users in the LGBT+ community, like [Glaad](#) and [Stonewall](#), to understand the lived experiences of users from this community in order to improve its policies and products to support that community.

Prioritisation:

- TikTok considers risks of dissemination of GBV content to be a **Tier 1** priority.
- This is due to the potential for online and real world physical, psychological and societal harms to potentially vulnerable individuals and the dynamic nature of this risk. TikTok will closely consider this prioritisation as more granular data becomes available (see 'Probability' above).
- Accordingly, in addition to continuing to proactively enforce its Community Guidelines policies on [Sexual Exploitation and Gender-Based Violence](#), TikTok's specialist teams will also remain highly vigilant to detect such risks and ensure preparedness to quickly mitigate and contain any emerging risks as they materialise.

Key further mitigation effectiveness improvements in line with Art. 35 of the DSA:

- **External engagement (Art. 35(1)(g)):** TikTok will engage with the organisations that are designated as "trusted flaggers" by EU member states to ensure efficient and priority

processing of their illegal content reports.

- **Media literacy (Art. 35(1)(i)):** TikTok will continue to roll-out resources for users affected by GBV to access its network of hotlines and community partners that can provide direct support and assistance to survivors, including across the EU.
- **Continued risk monitoring and vigilance (Art. 35(1)(f)):** TikTok will continue to devote cross-functional resources on a priority basis to its processes for handling risks associated with GBV Content. In addition, TikTok shall continue to collect and monitor relevant data as part of its transparency reporting obligations under the DSA.

Risk Mitigations - Table 4	Risks of gender-based violence content
TikTok's risk-mitigation measures in accordance with Art. 35(1) of the DSA (a) to (k), plus any other relevant measures:	
(a) Adaptation of feature or platform design, including online interfaces (as defined in Art. 3(m) DSA):	<p>TikTok implements a range of measures to prevent and mitigate the risk of the dissemination of GBV Content on or through the Platform, including through the following features, Platform design, and relevant technical safety measures:</p> <ul style="list-style-type: none"> • For all users, TikTok's default settings include a filter for spam and offensive comments in order to protect users from offensive or harmful comments; • Duet and Stitch features are default set to "off" and users have the choice to change these controls each time they post a video. This enables the creator to limit the risk that other users use these features to create GBV Content; • TikTok may, from time to time, block certain search results associated with GBV Content from appearing on the Platform; • Following external research on the topic, TikTok has implemented the "consider before you comment" prompt to remind users about its Community Guidelines and provide them the opportunity to edit comments before sharing with other users. This provides users with the opportunity to re-consider before they post content that potentially offensive or harmful comments;²⁵ and • TikTok has a range of measures in place to mitigate the risk of a livestream containing GBV Content.
(b) Adaption of terms and conditions(as defined in Art. 3(u) DSA) and their enforcement:	<p>TikTok prohibits GBV Content on the Platform through a combination of TikTok's Terms of Service and the Community Guidelines, and takes a range of measures to generate awareness and inform users. Under the heading of "Safety and Civility", TikTok's Community Guidelines prohibit Sexual Exploitation and Gender-Based Violence, stating that '<i>We do not allow sexual exploitation or gender-based violence, including non-consensual sexual acts, image-based sexual abuse, sextortion, physical abuse, and sexual harassment</i>'. The Community Guidelines provide further detail to</p>

²⁵ [Short of Suspension: How Suspension Warnings Can Reduce Hate Speech on Twitter](#), Yildirim, M., Nagler, J., Bonneau, R., & Tucker, J. (2021). *Perspectives on Politics*, 1-13.

	users about what each of these terms mean.
(c) Adoption of content moderation (as defined in Art. 3(t) DSA) processes:	Please see the 'Key Information about TikTok' section for a description of TikTok's content moderation processes.
(d) Testing and adaption of algorithmic systems, including recommender systems (as defined in Art. 3(s) DSA):	<p>TikTok's personalised FYF is one of the primary means through which users consume video content on the Platform.</p> <p>In addition to TikTok's content moderation policies and processes (see 'Key information about TikTok' above), further adaptations to mitigate the risk of violative content being recommended in the FYF are:</p> <ul style="list-style-type: none"> • TikTok detects and removes certain violative content at the point of attempted upload to the Platform or otherwise removes such content when detected. Such detected content cannot therefore be displayed on the FYF; • TikTok manually reviews video content when it reaches certain levels of popularity in terms of the number of video views, reducing the risk of violative content being shown in the FYF or otherwise being widely disseminated; and • TikTok offers users the tools to diversify the content displayed, to understand why videos have been recommended, to choose that certain keywords will not be displayed to them and to reset their FYF as if they were new to TikTok.
(e) Adaptation of advertising systems:	Please see the 'Key Information about TikTok' section for a description of TikTok's content moderation processes.
(f) Reinforcing risk detection measures:	<p>TikTok operates a number of specialist teams that deal with GBV in particular and who play a role in reinforcing risk detection measures:</p> <ul style="list-style-type: none"> • The Trust & Safety Product Policy team oversees the detailed policies for Violent Behaviours & Dangerous Actors and Harassment & Hateful Behaviour. The team includes experts who have undertaken specialist study and research, and who have deep sectoral experience in policy development in this field; and • The Trust & Safety team also includes an outreach function who are responsible for managing TikTok's relationships with existing partners and outreach initiatives and for developing new outreach partnerships. This work is extremely important to enable strategic collaborations with external experts, specialists and academics in the field of countering hateful behaviour to ensure that these teams and resulting policies are kept up to date on emerging developments in this space.
(g) Cooperation with trusted flaggers:	TikTok operates its Community Partner Channel, whereby onboarded NGOs perform a similar role to designated trusted flaggers (under the DSA) to submit reports of suspected harmful content, including suspected hate speech and forms of GBV Content. Those onboarded to the Community Partner Channel include various European and international NGOs who can report suspected harmful material using a dedicated channel to TikTok's

	<p>Incident Management team, who review such reports on a priority basis, including the following relevant to the risks of GBV Content:</p> <ul style="list-style-type: none"> • Arcigay (Italy); • BeLonG To Youth Services (Ireland); • The International Network Against Cyber Hate (“INACH”); • Internet Watch Foundation (“IWF”); • Linha Internet Segura (Safer Internet Helpline); and • Stopline (Austria). <p>Given the severity of harms related to GBV Content (and other illegal content types) TikTok has put in place a DSA Trusted Flagger Engagement Strategy so that it is prepared to work closely with newly designated trusted flaggers when they are designated.</p>
(h) Cooperation with other platforms through the codes of conduct/crisis protocols:	<p>TikTok signed the EU Code of Conduct on Hate Speech in September 2020. As part of this initiative, Code signatories are evaluated on a yearly basis by participating NGOs from the European region. These NGOs conduct a hate speech monitoring and reporting test over a four week period. They submit their results to the European Commission and the data is then shared individually with platforms who can review and discuss any anomalies with NGOs. In general, the tests seek to understand how quickly and reliably platforms respond to reports on hate speech content.</p>
(i) Awareness-raising measures for recipients of the services:	<p>TikTok undertakes the following measures:</p> <ul style="list-style-type: none"> • TikTok’s online Safety Centre webpage, has a page dedicated to Countering hate on TikTok; • TikTok’s online Safety Centre page on Sexual assault resources contains information and support for survivors of sexual assault contains links to various EU-based organisations with whom TikTok has developed partnerships, including: South West Grid for Learning (SWGfL), which runs the StopNCII.org service (and also runs the Revenge Porn Helpline in the UK); SOS Viol (Belgium); Danner (Denmark); Rape Crisis Centre Tukinainen (Finland); Dublin Rape Crisis Centre (Ireland); and Kvinnofridslinjen (Sweden). • TikTok’s Help Centre contains 'how to' explanations to allow users to learn about its content moderation practices and how to report violative content. • TikTok has also undertaken several recent on Platform media literacy campaigns such as: <ul style="list-style-type: none"> ◦ #SwipeOutHate which has been operated in conjunction with major sporting tournaments in Europe in 2022 and 2023; and ◦ #SaferTogether which has been operated to improve awareness and usage of TikTok's Safety, Privacy and Well-Being features.
(j) Targeted measures to protect the rights of the child:	<p>The following measures mitigate the impacts of Hate GBV content on minors:</p> <ul style="list-style-type: none"> • The options for commenting on videos posted by users aged 13-15 are restricted. These users can only choose to allow "Friends" or "No One" to comment on their Videos; they cannot choose to allow "Everyone" to comment. This reduces the risk of individuals being

	<p>able to share comments containing Hate Speech with users under 16; and</p> <ul style="list-style-type: none">• TikTok does not permit other users to Duet or Stitch with or download videos created by users aged 13-15. For users aged 16-17, the Duet and Stitch settings are set to 'Friends' by default.
(k) Measures to identify and address inauthentic content and behaviours:	<p>There is a risk that bad actors may intentionally manipulate the Platform or make inauthentic use of the Platform. TikTok takes a range of measures to prevent and mitigate the risk, including:</p> <ul style="list-style-type: none">• Account impersonation may enable hateful organisations and users to craft identities that appear trustworthy and relatable to their intended audience. Users are not allowed to use someone else's name, biographical details, or profile picture in a misleading manner. For example, impersonation may enable hateful organisations to craft identities that appear trustworthy and relatable to their intended audience. TikTok uses a range of methods to detect and remove these accounts; and• TikTok does not allow coordinated attempts to influence or sway public opinion while also misleading individuals, the community, or its systems about an account's identity, approximate location, relationships, popularity, or purpose. TikTok investigates and removes these operations, focusing on [REDACTED] to determine if actors are engaging in a coordinated effort to mislead TikTok's systems or its community.

9. RISKS OF TERRORIST CONTENT

Description of the risk:

- 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;
 - 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.

Key mitigation measures put in place:

- Risk Mitigations - Table 5 sets out a summary of risk mitigation measures in place with reference to Art. 35 (1) (a) to (k) of the DSA.
- Key mitigations for this risk are: (1) TikTok’s policies and proactive content moderation enforcement systems and processes; (2) the limits placed on searching for violative content; (3) the identification and removal from the Platform of recognised Terrorist groups; (4) incident management and risk detection activities including liaison with specialist law enforcement units; and (5) specific actions taken in relation to the For You recommender system.
- TikTok gives appropriate consideration to balancing freedom of expression when designing mitigation measures. However, it takes a zero tolerance approach in seeking to prevent violent extremists from using the Platform to instigate violence or spread harmful ideology.

- Risks relating to Terrorist Content tend to be highly localised due to regional, cultural and linguistic context in which they exist and which leads to differences across the EU in terms of how such speech is formulated and expressed and those who are targeted. An assessment of the relevant local/regional, cultural and linguistic factors is critical and is integrated as part of TikTok's Trust & Safety team's risk management processes, for example in maintaining a global risk calendar.

Key Data relied on:

- TikTok reported in its Q1 2023 Community Guidelines Enforcement Report that it detected and removed 1,305,534 pieces of video content globally under TikTok's violent extremist content policies and processes. In addition:
 - 94.9% was proactively detected by TikTok;
 - 77.4% was removed before there were any views of that content; and
 - 85.9% was removed within 24 hours of upload.
- TikTok published its first transparency report under the TCO Regulation in February 2023. This report assists in identifying the risk of Terrorist Content on the Platform during the relevant reporting period of June to December 2022. During that period, TikTok did not receive any such removal orders from competent authorities under the TCO Regulation. The Report also contains the number of removals of Terrorist Content in Germany, Austria and France (where local laws expressly require Platforms to report this action taken). 53,358 items of Terrorist Content were removed for violations of local laws relating to violent extremism.
- TikTok's transparency reports include information about any covert influence operations it identified and removed from the Platform. The Q1 2023 report does not contain any evidence of known terrorist groups engaging in covert coordinated activity on the Platform.

Severity:

- Terrorist content is criminal in nature, and can have very serious negative consequences for those exposed to it, as well as society at large. Being exposed to violent and graphic content, such as live footage of a terrorist attack or the aftermath may also cause distress or psychological harm, or lead to extreme discomfort.
- Terrorist content may also seek to promote extremism, radicalise and recruit followers, and to facilitate and direct terrorist activity, and/or to seek funding for such purposes. At the extreme end, it could involve threats to the life or physical safety of individuals, thus potentially causing major public safety/security risks.
- Where Terrorist Content involves incitement or encouragement, in certain circumstances (in particular for vulnerable viewers) it may contribute to the risk of viewers being radicalised and seeking to further support, incite, or participate in terrorism.
- TikTok's functionality makes it hard for users to discover specific types of material intentionally. TikTok therefore concludes that a bad actor seeking to recruit or radicalise would not choose TikTok over other more limited means of clandestine communication that are available.
- The duration and remediability of the impacts of the dissemination of Terrorist Content will depend on the specific nature of the content.
- TikTok's key data above demonstrates that TikTok's mitigation measures have a significant impact on the scale and potential duration of harm to users from Terrorist Content.
- TikTok assesses the potential severity of the risk of the dissemination of any Terrorist Content to be material due to: (1) the nature of the harms that may occur, including a risk to the physical and psychological state of individuals, as well as the potential for societal impact to

individual or several EU member states; and (2) TikTok's proactive efforts to enforce its policies, as demonstrated by the mitigation measures referred to above and in particular to TikTok's 94.9% proactive detection rate. In particular this assessment results from the measures put in place to restrict the scale of the dissemination of Terrorist Content, the duration for which it may circulate on the Platform, and the measures put in place to prevent repeated viewing of such content.

Probability:

- Terrorist Content is manifestly illegal content and, where it depicts violent extremism, is identifiable on its face and not appealing to broad audiences.
- TikTok understands and appreciates that such content can be created with the intention of shocking its audience. There is a risk that such content could be shared innocently as a warning or as counter speech. TikTok does not consider that its Platform functionality lends itself to the clandestine collaboration of, or recruitment to, terrorist groups.
- There could be a heightened risk of a livestream containing Terrorist Content before it is detected. However, no such incident has yet occurred on the Platform. TikTok has run a 'red team' simulation as part of its preparedness. Please see the Case Study entitled "Proactive mitigation measures relating to violent extremism risks" below.
- Given that it is highly criminal in nature, Terrorist Content is very unlikely to be popular content, such that it is selected to be recommended for users to view via the For You page, or content that has the potential to 'go viral' on TikTok.
- TikTok's Ad Policies are designed to protect users from fake, fraudulent, or misleading content. Advertiser accounts and ad content must comply with TikTok's Community Guidelines, which prohibits content disseminating support, praise or glorification of terrorist organisations and violent extremist organisations. TikTok's pre-moderation of ads means any advertising on the Platform is unlikely to include forms of Terrorist Content.
- TikTok strives to prevent the upload of, or otherwise remove, Terrorist Content from the Platform. TikTok's proactively enforces its policies on violent extremism, operates reporting channels for users, third parties and conducts ongoing risk detection activities including liaison with specialist law enforcement units. The key data above demonstrates that TikTok's mitigation measures have a significant impact on reducing the likelihood of dissemination of Terrorist Content on the Platform. TikTok considers that these measures demonstrate that it is unlikely that there is widespread dissemination of Terrorist Content on the Platform.

Key stakeholder engagement:

- TikTok is a member of [TechAgainstTerrorism](#), which is an important source of best practice in supporting the technology industry to tackle terrorist exploitation of the internet, whilst respecting human rights.
- TikTok is a member of the EU Internet Forum which gathers together EU member states, industry, academia, law enforcement, European agencies and international partners to discuss and address the challenges posed by the presence of malicious content online, including terrorist and violent extremist content. TikTok has committed to an EU Crisis Protocol – a rapid response initiative to contain the spread of terrorist and violent extremist content online.

Prioritisation:

- TikTok considers risks of dissemination of Terrorist Content to be a **Tier 2** priority.
- This is due to the potential severity of online psychological and societal harms. TikTok takes

this view notwithstanding its analysis that its functionality does not lend itself to the amplification of these risks. TikTok has nevertheless taken steps to prepare for such a scenario, as described in the Case Study in the next section.

- TikTok's specialist teams will closely monitor and remain highly vigilant to such risks, and will continue to proactively enforce its Community Guidelines policies on [Violent and Hateful Organisations and Individuals](#), as well as ensuring preparedness to quickly mitigate and contain any emerging risks as they materialise.

Key further mitigation effectiveness improvements in line with Art. 35 of the DSA:

- **Detection developments (Art. 35(1)(f)):** Following implementation of both internal and external processes to receive, review and action orders to act against illegal content (concerning Terrorist Content) under Art. 9 DSA, TikTok will continue to monitor the effectiveness of these processes. TikTok should also undertake outreach to engage with competent authorities across EU member states in relation to these processes and take account of any feedback received
- **External engagement (Art. 35 (1)(g)):** TikTok will engage with organisations when they are newly designated as "trusted flaggers" by EU member states (per Art. 22 DSA). TikTok will 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; and
- **Continued risk monitoring and vigilance (Art. 35(1)(f)):** TikTok will continue to devote cross-functional resources on a priority basis to its processes for handling Terrorist Content risks. In addition, TikTok shall continue to collect and monitor relevant data as part of its transparency reporting obligations under the DSA.

Risk Mitigations - Table 5	Risks of terrorist content
TikTok's risk-mitigation measures in accordance with Art. 35(1) of the DSA (a) to (k), plus any other relevant measures:	
(a) Adaptation of feature or platform design, including online interfaces (as defined in Art. 3(m) DSA):	<p>TikTok implements a range of measures to prevent and mitigate the risk of the dissemination of Terrorist Content on or through the Platform, including through the following features, Platform design, and relevant technical safety measures:</p> <ul style="list-style-type: none"> • TikTok may, from time to time, block certain search results associated with Terrorist Content from appearing on the Platform; and • TikTok has a range of measures in place to mitigate the risk of a livestream being capable of containing Terrorist Content.
(b) Adaption of terms and conditions(as defined in Art. 3(u) DSA) and their enforcement:	<p>TikTok very clearly prohibits Terrorist Content on the Platform through a combination of TikTok's Terms of Service and the Community Guidelines, and takes a range of measures to generate awareness and inform users.</p> <p>TikTok's Community Guidelines state that "<i>We do not allow the presence of violent and hateful organisations or individuals on our platform</i>". The</p>

	<p>Community Guidelines also detail that TikTok prohibits content disseminating support, praise or glorification of terrorist organisations and violent extremist organisations.</p> <p>TikTok's Trust & Safety teams define and operate processes to determine if a group is a terrorist group, so that it can implement effective detection and moderation strategies (this involves identifying terrorist groups, associated individuals, symbols, slogans, etc). This work is informed by resources such as the UN and EU terrorist lists. TikTok may also take off-Platform behaviour into consideration as it enforces its policies, such as an account belonging to the leader of a known hate group, to protect people against harm.</p>
(c) Adoption of content moderation (as defined in Art. 3(t) DSA) processes:	<p>Please see the 'Key Information about TikTok' section for a description of TikTok's content moderation processes.</p> <p>In addition, in order to maintain its detection models, TikTok works with various external parties to keep its keyword lists relevant to Terrorist Content up to date. Similar strategies are deployed to detect violating text content in user profile bios or username handles, which may be indicative of accounts associated with Terrorist Content. TikTok maintains a "blacklist" of hyperlinks, which have been found to be associated with previously detected Terrorist Content, so that it can detect, disrupt and remove such links that lead to off-Platform content.</p>
(d) Testing and adaption of algorithmic systems, including recommender systems (as defined in Art. 3(s) DSA):	<p>TikTok's personalised FYF is one of the primary means through which users consume video content on the Platform.</p> <p>Relevant adaptations are:</p> <ul style="list-style-type: none"> • TikTok detects and removes certain Terrorist Content at the point of attempted upload to the Platform or otherwise removes such content when detected. Such detected content cannot therefore be displayed on the FYF; • TikTok manually reviews video content when it reaches certain levels of popularity in terms of the number of video views, reducing the risk of violative content being shown in the FYF or otherwise being widely disseminated; and • TikTok offers users the tools to understand why videos have been recommended, to diversify the content displayed, to choose that certain keywords will not be displayed to them and to reset their FYF as if they were new to TikTok.
(e) Adaptation of advertising systems:	<p>Please see the 'Key Information about TikTok' section for a description of TikTok's content moderation processes.</p>
(f) Reinforcing risk detection measures:	<p>TikTok operates a number of specialist teams that deal with illegal content and Terrorist Content in particular and who play a role in reinforcing risk detection measures:</p> <ul style="list-style-type: none"> • The Incident Management team is a multidisciplinary team of experts who work together to identify, detect, and mitigate risk in response to escalation scenarios. This team provides 24/7 incident management and risk handling coverage and support, which primarily involves removal of Terrorist Content from the Platform;

	<ul style="list-style-type: none"> • The Law Enforcement Response team has responsibility for reviewing law enforcement and governmental data disclosure requests. Where such requests concern suspected terrorist-related offences, these are treated as a top priority; • The Law Enforcement Outreach team frequently engage in outreach with national law enforcement authorities across the EU and with international agencies (such as Europol and Interpol, and with trusted law enforcement agencies outside the EU) to establish working relationships to facilitate clear and effective reporting and other channels of communication. This team also use this outreach as part of TikTok's risk detection and intelligence gathering on new and emergent risks, including as they relate to Terrorist Content; and • TikTok's Emergency Response team provides 24/7 coverage by experienced safety specialists to handle internal escalations of suspected emergency situations and for handling emergency data disclosure requests submitted by law enforcement, which may involve imminent terrorist-related risks.
(g) Cooperation with trusted flaggers:	<p>TikTok already operates its Community Partner Channel, through which onboarded NGOs perform a similar role to designated trusted flaggers (under the DSA) to submit reports of suspected harmful content. Many groups have been on-boarded as part of this programme, including the following relevant to the risks of Terrorist Content:</p> <ul style="list-style-type: none"> • Pharos (France's cyber police agency); • Violence Prevention Network (Germany); and • Portugal's Safer Internet Helpline. <p>Given the severity of harms related to Terrorist Content (and other illegal content types) TikTok has put in place a DSA Trusted Flagger Engagement Strategy so that it is prepared to work closely with newly designated trusted flaggers when they are designated.</p>
(h) Cooperation with other platforms through the codes of conduct/crisis protocols:	<p>TikTok's internal risk detection and monitoring capabilities are supplemented by its partnerships with leading non-governmental counterterrorism organisations specialising in tracking and analysing online activity of the global violent extremist community.</p> <p>These partners support technology platforms aiming to eradicate terrorist and violent extremist content from their platforms and have provided governments and institutions worldwide with verified, actionable intelligence and analysis on designated terrorist and violent extremist groups. Its partners' services include monitoring, crisis and incident coverage, the provision of databases of Terrorist Content for the purposes of comparison and threat intelligence gathering.</p>
(i) Awareness-raising measures for recipients of the services:	<p>TikTok undertakes the following measures:</p> <ul style="list-style-type: none"> • TikTok's Transparency Centre has dedicated content explaining 'Our approach to content moderation' and 'Combating hate and violent extremism'; and • TikTok's Help Centre contains 'how to' explanations to allow users to learn about its content moderation practices and how to report

	violative content.
(j) Targeted measures to protect the rights of the child:	TikTok does not take any specific mitigation measures in relation to the risk that minors may experience Terrorist Content. This is because TikTok takes a zero tolerance approach to all illegal content and seeks to remove it before any user, of any age, is exposed to it.
(k) Measures to identify and address inauthentic content and behaviours:	<p>There is a risk that bad actors may intentionally manipulate the Platform or make inauthentic use of the Platform. TikTok takes a range of measures to prevent and mitigate the risk, including:</p> <ul style="list-style-type: none"> Account impersonation may enable terrorist organisations to craft identities that appear trustworthy and relatable to their intended audience. TikTok's users are not allowed to use someone else's name, biographical details, or profile picture in a misleading manner. For example, impersonation may enable terrorist organisations to craft identities that appear trustworthy and relatable to their intended audience. TikTok uses a range of methods to detect and remove these accounts; and TikTok does not allow coordinated attempts to influence or sway public opinion while also misleading individuals, its community, or its systems about an account's identity, approximate location, relationships, popularity, or purpose. TikTok investigates and removes these operations, focusing on [REDACTED] to determine if actors are engaging in a coordinated effort to mislead TikTok's systems or its community.

Case Study: Proactive mitigation measures relating to violent extremism risks

Context: In recent years multiple high profile terrorist and violent extremist attacks in Europe and elsewhere have been livestreamed in real time on online platforms. While no such incident has occurred on TikTok (where TikTok is the primary or first source for dissemination), TikTok remains highly vigilant to the risk it could be a primary or secondary source for such content. This Case Study explains some of the proactive risk mitigation measures TikTok operates to ensure its readiness to prevent and contain such risks.

Readiness-testing through risk simulation: In September 2022 TikTok conducted a simulation exercise of a live shooting incident on the Platform. The purpose was to assess readiness for handling such scenarios and identify areas for improvement. Adapted to imitate a real-world event, this simulation exercise comprised various simulated "moves" over several hours with teams fed information on an iterative basis. The exercise involved over 40 stakeholders from various internal teams and was designed to provide an opportunity for cross functional stakeholders to test their reaction, processes and team collaboration during potential livestreamed attacks on TikTok. The exercise tested the participants' ability to adapt to the evolving scenario and the functioning of existing strategies and processes for engagement with internal and external stakeholders.

Outputs of the simulation exercise: TikTok's work in ensuring readiness to prevent and contain risks relating to violent extremism are informed by its extensive engagement with external experts, including across the academic, public policy, civil society and law enforcement communities. TikTok's Risk Assessment & Policy teams consulted with external experts (including academic experts who study livestreamed terrorist attacks) in the development of, and preparation for, the simulation in order to better understand the nature of such risks and to ensure the simulation was reflective of a real life scenario. TikTok presented its proposed end-to-end approach to addressing mass casualty incidents, should they ever occur on the Platform (including its areas for continuous improvement - see below)) to those partners for their review. Feedback from those experts validated that TikTok has an industry leading, robust and informed end-to-end approach.

Continuous improvement: In line with TikTok's commitment to continuous improvement in its risk management practices, TikTok has taken a number of actions in response to the exercise to bolster its readiness and incident management processes, to ensure it is able to respond quickly and effectively. These include:

- TikTok has implemented a specific policy to handle real-time mass casualty incidents. TikTok has also trained its moderation teams to respond to content of this nature (particularly content that appears in the aftermath of such attacks, such as reproduced content);
- TikTok has augmented its internal risk detection and monitoring capabilities with supplementary information from partnerships with external threat detection partners, who assist by providing real time insights and information on detecting credible threats for escalation;
- TikTok has designed its processes in response to any livestream of a mass-casualty event, which includes imposing search interventions (such as blocking words commonly used to search for violative content, as well other restrictions on users ability to spread violative content more widely) and a safety centre to prevent re-victimisation of content posted in the aftermath; and
- TikTok has designed its post-detection response to ensure it is user-centric, in conjunction with regional wellness organisations that focus on mental health. This includes: (i) directing users to appropriate resources to take care of their mental and emotional wellbeing; and (ii) increasing awareness and condemning acts of violence to decrease radicalisation.

10. RISKS OF ILLEGAL HATE SPEECH CONTENT

Description of the risk:

- 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, Art. 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;
 - 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 that Art. 21 of the Charter.

Key mitigation measures put in place:

- Risk Mitigations - Table 6 sets out a summary of risk mitigation measures in place with reference to Art. 35 (1) (a) to (k) of the DSA.
- Key mitigations for this risk are: (1) TikTok’s policies and proactive content moderation enforcement systems and processes; (2) the limits placed on searching for violative content; (3) the identification and removal from the Platform of hateful organisations, promotion or material support of hateful actors and hateful ideologies; (4) incident management and risk detection activities including liaison with specialist law enforcement units; and (5) specific actions taken in relation to the For You recommender system.
- TikTok gives appropriate consideration to balancing freedom of expression when designing mitigation measures. However, it takes significant steps to prevent Hate Speech and hateful behaviours on the Platform.
- As risks relating to Hate Speech tend to be highly localised within Europe, through its Trust & Safety regional policy teams, TikTok places an emphasis on ensuring it takes a localised approach in moderating Hate Speech that is informed by the relevant language, culture and context.

Key data relied on:

- TikTok reported in its Q1 2023 Community Guidelines Enforcement Report that it detected and removed 2,297,184 pieces of video content globally under TikTok's hateful behaviour content policies and processes. In addition:
 - 89.2% of such content was proactively detected by TikTok;
 - 74.9% was removed before there were any views of that content; and
 - 83.8% was removed within 24 hours of upload.
- In its most [recent evaluation](#) of the EU Code of Conduct on Countering Illegal Hate Speech Online (for 2022), the published results indicate the increasing effectiveness of TikTok's approach in tackling Hate Speech. TikTok improved its time of assessment (i.e., TikTok's time of assessment within 24-hours increased from 82.5% in 2021 to 91.7% in 2022, compared to the overall average of 64.4%).²⁶ Large-scale events or celebrations can see a spike in hateful online behaviour, so TikTok's teams take this into account as they proactively identify and scenario-plan for such risks.

Severity:

- The spread of illegal Hate Speech online has substantial negative effects for the groups or individuals that it targets. Hate Speech also negatively impacts society more generally by undermining democratic principles of freedom of expression, tolerance and non-discrimination, potentially through indoctrination of hateful ideologies and discriminatory views. Those who are the subject of, or inadvertently exposed to, Hate Speech, may suffer distress.
- Risks relating to Hate Speech tend to be highly localised due to the regional, cultural and linguistic context in which they exist and which leads to differences across Europe in terms of how such speech is formulated and expressed and those who are targeted. There is a very real practical challenge in effectively moderating Hate Speech dynamically and at scale across a region such as Europe with such a high level of cultural, linguistic and regional diversity in a reasonable manner without disproportionately impacting on users' freedom of expression.
- The duration and remediability of the impacts of the dissemination of Hate Speech will depend on the specific nature of the content.
- TikTok's key data above demonstrates that TikTok's mitigation measures have a significant impact on the scale and potential duration of harm to users from Hate Speech.
- TikTok assesses the potential severity of the risk of dissemination of Hate Speech to be moderate, due to: (1) the nature of the harm involving negative effects to individuals – such as distress – as well as society through its potential impact on the principles of freedom of expression, tolerance, and non-discrimination; and (2) TikTok's proactive efforts to enforce its policies, as demonstrated by the mitigation measures referred to above and in particular to TikTok's 89.2% proactive detection rate. In particular this assessment results from the measures put in place to restrict the scale of the dissemination of Hate Speech, the duration for which it may circulate on the Platform, and the measures put in place to prevent repeated viewing of such content.

²⁶ EU Code of Conduct against online hate speech: latest evaluation, Nov 2022. Noting that the EU Code of Conduct on Countering Illegal Hate Speech Online pre-dates the DSA, as codes of conduct (under Art. 45 of the DSA) and crisis protocols (under Art. 48 of the DSA) have not yet been drawn up in the context of the DSA (and that the European Board for Digital Services has not yet been convened).

Probability:

- TikTok strives to prevent the upload of, or otherwise remove, Hate Speech from the Platform. TikTok proactively enforces its policies on hateful behaviour and operates reporting channels for users.
- The key data above demonstrates that TikTok's mitigation measures have a significant impact on reducing the likelihood of dissemination of Hate Speech on the Platform.
- TikTok considers that it is possible that there will remain some level of dissemination of Hate Speech on the Platform, whilst recognising that the definition and incidence of Hate Speech necessarily covers a wide span of behaviours and risks.
- Under TikTok's Ad Policies, Advertiser accounts and ad content must comply with TikTok's Community Guidelines, which prohibits Hate Speech. TikTok's pre-moderation of ads means any advertising on the Platform is unlikely to include forms of Hate Speech.

Key stakeholder engagement:

- TikTok is a committed signatory to the [EU Code of Conduct on Countering Illegal Hate Speech Online](#), and actively engages with the European Commission and other stakeholders.
- TikTok continues its engagements with stakeholders who are committed to combating online hate speech, including the [INACH Network](#), the World Jewish Congress, the [Glaad](#) and [Stonewall](#), and [European Observatory of Online Hate](#).

Prioritisation:

- TikTok considers risks of dissemination of Hate Speech to be a **Tier 2** priority.
- This is due to the potential for online physical, psychological and societal harms and the dynamic nature of this risk.
- TikTok's specialist teams will continue to closely monitor and remain vigilant of such risks, which remain highly dynamic, and will continue to proactively enforce its Community Guidelines policies on [Hate Speech and Hateful Behaviours](#).

Key further mitigation effectiveness improvements in line with Art. 35 of the DSA:

- **Detection developments (Art. 35(1)(f)):** TikTok will plan to expand its collaboration with external partners in order to develop enhanced intelligence-gathering capacities in relation to Hate Speech. In particular, this should assist in early detection of emerging and new forms of Hate Speech.
- **External engagement (Art. 35(1)(g)):** TikTok will engage with the organisations that are designated as "trusted flaggers" by EU member states to ensure efficient and priority processing of their illegal content reports. TikTok is a committed signatory of the EU Code of Conduct on Hate Speech, and will continue its engagement with the European Commission and key stakeholders as part of this important initiative to combat online Hate Speech.
- **Media literacy (Art. 35(1)(i)):** Building on the success of media literacy campaigns, such as the #SwipeOutHate campaigns, TikTok should continue such initiatives and consider further media literacy measures to generate awareness of issues relating to illegal content and of the available safety tools.
- **Continued risk monitoring and vigilance (Art. 35(1)(f)):** TikTok will continue to devote cross-functional resources on a priority basis to its processes for handling Terrorist Content risks. In addition, TikTok shall continue to collect and monitor relevant data as part of its transparency reporting obligations under the DSA.

Risk Mitigations - Table 6	Risks of illegal hate speech
TikTok's risk-mitigation measures in accordance with Art. 35 (1) of the DSA (a) to (k), plus any other relevant measures:	
(a) Adaptation of feature or platform design, including online interfaces (as defined in Art. 3(m) DSA):	<p>TikTok implements a range of measures to prevent and mitigate the risk of the dissemination of Hate Speech on or through the Platform, including through the following features, Platform design, and relevant technical safety measures:</p> <ul style="list-style-type: none"> • TikTok has implemented various features that act to limit the publication of violative content. For all users, TikTok's default settings include a filter for spam and offensive comments in order to protect users from offensive or harmful comments; • Duet and Stitch features are default set to "off" and users have the choice to change these controls each time they post a video. This enables the creator to limit the risk that other users use these features to create content that includes Hate Speech; • TikTok may, from time to time, block certain search results associated with Hate Speech from appearing on the Platform; • Following external research on the topic, TikTok has implemented the "consider before you comment" prompt to remind users about the Community Guidelines and provide them the opportunity to edit comments before sharing with other users. This provides users with the opportunity to re-consider before they post content that potentially offensive or harmful comments²⁷; and • TikTok has a range of measures in place to mitigate the risk of a livestream containing Hate Speech.
(b) Adaption of terms and conditions(as defined in Art. 3(u) DSA) and their enforcement:	<p>TikTok very clearly prohibits Hate Speech on the Platform through a combination of TikTok's Terms of Service and the Community Guidelines, and takes a range of measures to generate awareness and inform users. Under the heading of "Safety and Civility", TikTok's Community Guidelines prohibit Hate Speech, stating that '<i>We do not allow any hateful behaviour, hate speech, or promotion of hateful ideologies</i>'. The Guidelines provide further detail to users about what each of these terms mean.</p>
(c) Adoption of content moderation (as defined in Art. 3(t) DSA) processes:	<p>Please see the 'Key Information about TikTok' section for a description of TikTok's content moderation processes.</p> <p>In addition, TikTok's Trust & Safety team define and operate the processes for determining a hateful ideology or whether a group is a hate organisation, which is important in order to implement effective detection and moderation strategies (and involves identifying organised hate groups, associated individuals, symbols, slogans, etc.). Such processes take account of various factors across a range of hateful ideologies including: fascism; white supremacy or</p>

²⁷ [Short of Suspension: How Suspension Warnings Can Reduce Hate Speech on Twitter](#), Yildirim, M., Nagler, J., Bonneau, R., & Tucker, J. (2021). *Perspectives on Politics*, 1-13.

	nationalism; and male supremacy (including incel). TikTok may also take off-Platform behaviour into consideration.
(d) Testing and adaption of algorithmic systems, including recommender systems (as defined in Art. 3(s) DSA):	<p>TikTok's personalised FYF is one of the primary means through which users consume video content on the Platform.</p> <p>In addition to TikTok's content moderation policies and processes (see 'Key information about TikTok' above), further adaptations to mitigate the risk of violative content being recommended in the FYF are:</p> <ul style="list-style-type: none"> • TikTok detects and removes certain violative content at the point of attempted upload to the Platform or otherwise removes such content when detected. Such detected content cannot therefore be displayed on the FYF; • TikTok manually reviews video content when it reaches certain levels of popularity in terms of the number of video views, reducing the risk of violative content being shown in the FYF or otherwise being widely disseminated; and • TikTok offers users the tools to diversify the content displayed, to understand why videos have been recommended, to choose that certain keywords will not be displayed to them and to reset their FYF as if they were new to TikTok.
(e) Adaptation of advertising systems:	Please see the 'Key Information about TikTok' section for a description of TikTok's content moderation processes.
(f) Reinforcing risk detection measures:	<p>TikTok operates a number of specialist teams that deal with Hate Speech in particular and who play a role in reinforcing risk detection measures:</p> <ul style="list-style-type: none"> • The Trust & Safety Product Policy team oversees the detailed policies for Violent Behaviours & Dangerous Actors and Harassment & Hateful Behaviour. The team includes experts who have undertaken specialist study and research, and who have deep sectoral experience in policy development in this field; and • The Trust & Safety team also includes an outreach function who are responsible for managing TikTok's relationships with existing partners and outreach initiatives and for developing new outreach partnerships. This work is extremely important to enable strategic collaborations with external experts, specialists and academics in the field of countering hateful behaviour to ensure that these teams and resulting policies are kept up to date on emerging developments in this space.
(g) Cooperation with trusted flaggers:	<p>TikTok already operates its Community Partner Channel, through which onboarded NGOs perform a similar role to designated trusted flaggers (under the DSA) to submit reports of suspected harmful content.</p> <p>In the context of the EU Code of Conduct on Countering Illegal Hate Speech Online, TikTok has onboarded over 30 EU-based NGOs and other organisations who monitor and report suspected Hate Speech to TikTok from EU member states.</p> <p>Given the severity of harms related to Hate Speech content (and other illegal content types) TikTok has put in place a DSA Trusted Flagger Engagement Strategy so that it is prepared to work closely with newly designated trusted flaggers when they are designated.</p>

<p>(h) Cooperation with other platforms through the codes of conduct/crisis protocols:</p>	<p>TikTok signed the EU Code of Conduct on Hate Speech in September 2020. As part of this initiative, Code signatories are evaluated on a yearly basis by participating NGOs from the European region. These NGOs conduct a Hate Speech monitoring and reporting test over a four week period. They submit their results to the European Commission and the data is then shared individually with platforms who can review and discuss any anomalies with NGOs. In general, the tests seek to understand how quickly and reliably platforms respond to reports on hate speech content.</p>
<p>(i) Awareness-raising measures for recipients of the services:</p>	<p>TikTok undertakes the following measures:</p> <ul style="list-style-type: none"> • TikTok's online Safety Centre webpage, has a page dedicated to Countering hate on TikTok; • TikTok's online Transparency Centre webpage, in the Our Commitments page, contains articles explaining TikTok's approach to <i>Keeping People Safe</i>, which includes separate articles that explains Our approach to content moderation and Combating hate and violent extremism; • TikTok's Help Centre contains 'how to' explanations to allow users to learn about its content moderation practices and how to report violative content; and • TikTok has also undertaken several recent on Platform media literacy campaigns such as: <ul style="list-style-type: none"> ◦ #SwipeOutHate which has been operated in conjunction with major sporting tournaments in Europe in 2022 and 2023; and ◦ #SaferTogether which has been operated to improve awareness and usage of TikTok's Safety, Privacy and Well-Being features; and ◦ An ongoing initiative to provide authoritative Information on the Holocaust when users search for related words, which TikTok delivered in partnership with the World Jewish Congress and UNESCO.
<p>(j) Targeted measures to protect the rights of the child:</p>	<p>The following measures mitigate the impacts of Hate Speech content on minors:</p> <ul style="list-style-type: none"> • The options for commenting on videos posted by users aged 13-15 are restricted. These users can only choose to allow "Friends" or "No One" to comment on their Videos; they cannot choose to allow "Everyone" to comment. This reduces the risk of individuals being able to share comments containing Hate Speech with users under 16; and • TikTok does not permit other users to Duet or Stitch with or download videos created by users aged 13-15. For users aged 16-17, the Duet and Stitch settings are set to 'Friends' by default.
<p>(k) Measures to identify and address inauthentic content and behaviours</p>	<p>There is a risk that bad actors may intentionally manipulate the Platform or make inauthentic use of the Platform. TikTok takes a range of measures to prevent and mitigate the risk, including:</p> <ul style="list-style-type: none"> • Account impersonation may enable hateful organisations and users to craft identities that appear trustworthy and relatable to their intended audience. Users are not allowed to use someone else's name, biographical details, or profile picture in a misleading manner. For example, impersonation may enable hateful organisations to craft identities that appear trustworthy and relatable to their intended

	<p>audience. TikTok uses a range of methods to detect and remove these accounts; and</p> <ul style="list-style-type: none">• TikTok does not allow coordinated attempts to influence or sway public opinion while also misleading individuals, the community, or its systems about an account's identity, approximate location, relationships, popularity, or purpose. TikTok investigates and removes these operations, focusing on [REDACTED] to determine if actors are engaging in a coordinated effort to mislead TikTok's systems or its community.
--	---

11. RISKS TO PUBLIC HEALTH FROM MEDICAL MISINFORMATION CONTENT

Description of the risk:

- 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);
 - 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 Art. 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.

Key mitigation measures put in place:

- Risk Mitigations - Table 7 sets out a summary of risk mitigation measures in place with reference to Art. 35 (1) (a) to (k) of the DSA.
- Key mitigations for this risk are: (1) TikTok's policies and proactive content moderation enforcement systems and processes; (2) TikTok's Fact-Checking Programme, with [IFCN-certified](#) independent fact-checking partners across Europe who assist TikTok in verifying, labelling or removing content; (3) in-app interventions for users to direct them to authoritative content; and (4) monitoring of health trends to ensure preparedness.
- As risks relating to Medical Misinformation can be, but are not necessarily, localised, TikTok's Fact-checking programme makes sure that any differences are understood and factored into planning and execution of risk mitigation strategies.

Key data relied on:

- TikTok does not separately report on content detected and removed from the Platform due to Medical Misinformation and content removed for these reasons is contained within its reporting on harmful misinformation/content.
- TikTok reported in its Q1 2023 Community Guidelines Enforcement Report that it detected and removed 908,927 pieces of video content globally under TikTok's policies and processes for integrity and authenticity (which comprises harmful misinformation/content as well as spam and fake engagement) In addition:
 - 94.8% of such content was detected and removed proactively, before any reporting;
 - 72.8% was removed before there were any views of that content; and
 - 76.6% was removed within 24 hours of upload.
- TikTok's reporting under the [CoPD](#) contains relevant information and metrics on its efforts to combat disinformation.
 - In Q1 & Q2 2023, fewer than 2 in 10,000 views occurred on content identified and removed for violating its policies on harmful misinformation/content (all, not just Medical Misinformation); i.e., across Europe, 140,635 videos were removed, with the number of views of such videos in Europe being 1,012,020,899 (the equivalent numbers in the EEA are 142,711 and 1,019,752,855 respectively).²⁸
 - The number of ads removed for violating TikTok's political content ad policy was 390 in the EU/395 in the EEA (for the 6-month period of 1 January to 30 June 2023, as indicated in its CoPD Report).²⁹

Severity:

- Regardless of intent, TikTok prohibits inaccurate, misleading, or false content that may cause significant harm to individuals or society. The spread of Medical Misinformation can mislead individuals and could potentially contribute to negative impacts on health and physical wellbeing, which may, which may, in extreme scenarios, have long-term health implications that are likely very difficult to reverse.
- Medical Misinformation risks can be localised, in particular with linguistic differences across Europe based on the language(s) spoken by relevant populations and other cultural nuances, but may also have a European-wide scale.
- Medical Misinformation risks can generally be remediated by implementing effective measures to proactively detect and remove medical misinformation, through fact-checking and by labelling such content, as well as reducing its negative impact through in-app interventions to lead users to reliable information.
- TikTok's key data above demonstrates that TikTok's mitigation measures have a significant impact on the scale and potential duration of harm to users from harmful misinformation/content.
- TikTok assesses the potential severity of risks to public health caused by Medical Misinformation on the Platform to be moderate, due to: (1) the potential for serious and potentially long-term negative impacts on health and physical wellbeing for those that act in reliance on harmful misinformation/content, such harms which may be difficult to reverse; and (2) TikTok's proactive efforts to enforce its policies, as demonstrated by the mitigation measures referred to above and in particular to TikTok's 94.8% proactive detection rate. In particular this assessment results from the measures put in place to reduce its negative impact through in-app interventions and media literacy resources that lead users to reliable information sources on medical and public health matters to counteract the potential negative effects of Medical Misinformation.

²⁸ For the period 1 January - 30 June 2023, as reflected in TikTok's Code of Practice on Disinformation, p. 135 - 137.

²⁹ For the period 1 January - 30 June 2023, as reflected in TikTok's Code of Practice on Disinformation, p.4 - 6.

Probability:

- Medical misinformation risks are by nature inherently more unpredictable, and new forms of medical misinformation can evolve quickly without advance warning. This requires an emphasis on risk monitoring and incident management capacities to quickly detect, contain and mitigate the severity of such evolving risks.
- The World Health Organisation stated in May 2023 COVID-19 was no longer a global health emergency. TikTok noted in its Q1 Code of Practice on Disinformation report that it had already seen a downward trend in its metrics in line with the de-escalation of the risk. However, it is possible that further global pandemics will occur in future.
- TikTok strives to prevent the upload of, or otherwise remove, Medical Misinformation from the Platform. TikTok assesses that it is possible that there will remain some level of Medical Misinformation on the Platform. However, the end to the COVID-19 pandemic reduces the immediate probability of this risk manifesting on the platform in high volumes. TikTok's teams will remain vigilant to detect, contain and mitigate any new Medical Misinformation risks as they arise.
- TikTok's Ad Policies are designed to protect users from fake, fraudulent, or misleading content. TikTok's Community Guidelines and Ad Policies prohibit Medical Misinformation. TikTok's pre-moderation of ads means any advertising on the Platform is unlikely to include forms of Medical Misinformation.

Key stakeholder engagement:

- TikTok's Fact-checking Programme includes a range of measures which help to identify emerging misinformation-related risks (including Medical Misinformation), including ongoing updating of its internal repository of fact-checked misinformation; policy consulting/trends analysis from its fact-checking partners; and insights and market expertise from its internal cross-functional teams on key issues. As part of its Fact-Checking Programme, TikTok works closely with its [IFCN-certified](#) independent fact-checking partners (TikTok's European fact-checkers include Agence France-Presse (AFP), Deutsche Presse-Agentur (dpa), Facta, Logically, Lead Stories, Newtral, Science Feedback, Teyit, and Reuters).
- TikTok has worked closely with the [World Health Organisation](#), various national European health authorities/ministries, and partners across Europe in connection with its media literacy and awareness campaigns to combat Medical Misinformation.

Prioritisation:

- TikTok considers risks to public health from the dissemination of Medical Misinformation to be a **Tier 2** priority.
- This is due to the World Health Organisation [declaring](#) the end of COVID-19 as a global health emergency in May 2023, which TikTok considers to reduce the speed at which the risk evolves.
- TikTok will nonetheless closely monitor the continued evolution of Medical Misinformation risks, and continue to enforce its Community Guidelines policies on [Misinformation](#) and maintain vigilance to deploy crisis response mechanisms to respond to any emerging health emergency.

Key further mitigation effectiveness improvements in line with Art. 35 of the DSA:

- **Detection improvements (Art. 35(1)(f)):** TikTok will further expand its Fact-Checking Programme by on-boarding new European-based fact-checking partners and increasing its

- operational coverage in Europe; and
- **Continued risk monitoring and vigilance (Art. 35(1)(f)):** TikTok will continue to devote cross-functional resources on a priority basis to its processes for handling Medical Misinformation risks.

Risk Mitigations - Table 7	Risks to public health from medical misinformation content
TikTok's risk-mitigation measures in accordance with Art. 35(1) of the DSA (a) to (k), plus any other relevant measures:	
(a) Adaptation of feature or platform design, including online interfaces (as defined in Art. 3(m) DSA):	<p>TikTok implements a range of measures to prevent and mitigate the risk of the dissemination of violative content on or through the Platform, including through the following features, Platform design, and relevant technical safety measures:</p> <ul style="list-style-type: none"> • TikTok 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 Medical Misinformation or not. The content is ineligible for the FYF and search restrictions are also applied; • TikTok will prompt users to (re)consider unverified content before sharing it to other users on the Platform; • TikTok has a range of measures in place to mitigate the risk of a livestream being capable of containing Medical Misinformation.
(b) Adaption of terms and conditions(as defined in Art. 3(u) DSA) and their enforcement:	<p>TikTok prohibits Medical Misinformation on the Platform through a combination of TikTok's Terms of Service and the Community Guidelines, and takes a range of measures to generate awareness and inform users. Under the heading of "Integrity and Authenticity", TikTok's Community Guidelines prohibit Medical Misinformation, stating that "Not Allowed: [...] Medical misinformation, such as misleading statements about vaccines, inaccurate medical advice that discourages people from getting appropriate medical care for a life-threatening disease, and other misinformation that poses a risk to public health."</p>
(c) Adoption of content moderation (as defined in Art. 3(t) DSA) processes:	<p>Please see the 'Key Information about TikTok' section for a description of TikTok's content moderation processes.</p>
(d) Testing and adaption of algorithmic systems, including recommender systems (as defined in Art. 3(s) DSA):	<p>TikTok's personalised FYF is one of the primary means through which users consume video content on the Platform.</p> <p>In addition to TikTok's content moderation policies and processes (see 'Key information about TikTok' above), additional adaptations to mitigate the risk of violative content being recommended in the FYF are:</p> <ul style="list-style-type: none"> • TikTok detects content that may be Medical Misinformation and if it cannot be determined by a fact checker whether it is/not such content,

	<p>TikTok may label it as ‘unverified content’. The content will then be ineligible for the FYF;</p> <ul style="list-style-type: none"> • TikTok detects and removes certain violative content at the point of attempted upload to the Platform or otherwise removes such content when detected. Such detected content cannot therefore be displayed on the FYF; • TikTok manually reviews video content when it reaches certain levels of popularity in terms of the number of video views, reducing the risk of violative content being shown in the FYF or otherwise being widely disseminated; and • TikTok offers users the tools to diversify the content displayed, to understand why videos have been recommended, to choose that certain keywords will not be displayed to them and to reset their FYF as if they were new to TikTok.
(e) Adaptation of advertising systems:	<p>Please see the ‘Key Information about TikTok’ section for a description of TikTok’s content moderation processes.</p> <p>In addition, with specific applicability to Medical Misinformation:</p> <ul style="list-style-type: none"> • TikTok’s Medical Misinformation policy expressly prohibits verifiably false medical advice that may cause negative health effects on an individual’s life. This includes a prohibition on information regarding natural or homoeopathic cures that may discourage people from seeking appropriate medical treatment in response to a life threatening disease. • TikTok’s ads policies (which apply in addition to the Community Guidelines) contain specific prohibitions on Medical Misinformation related to COVID-19 and also make it clear that TikTok prohibits ads for alcohol, weight control/management, high fat, salt and sugar goods, prescription and non-prescription drugs and tobacco and related products. In addition, TikTok prohibits ads containing: <ul style="list-style-type: none"> • misleading claims about weight loss or cures for incurable diseases, alcohol, weight control/management; • misleading claims or inconsistent information, which could include content posing risks to public health, e.g., misleading claims about weight loss or cures for incurable diseases; • sensational or shocking content, which could include content posing risks to public health, e.g., content that depicts or promotes ingesting substances that are not meant for consumption or could lead to severe harm; and • content that could incite fear or panic, e.g., content that depicts or promotes phrases such as “you may be in danger” or “virus detected, remove it now”.
(f) Reinforcing risk detection measures:	<p>Please see the corresponding section in the Election Misinformation section of this Report for details of TikTok’s Fact-checking Programme.</p>
(g) Cooperation with trusted flaggers:	<p>Please see the corresponding section in the Election Misinformation section of this Report for details.</p>
(h) Cooperation with other platforms through the codes	<p>TikTok is a signatory to the CoPD and is a participant in the Permanent Taskforce and subgroups set up under the CoPD. TikTok has produced two reports from the period of Q4 2022 and Q1 & Q2 2023 including a Crisis</p>

of conduct/crisis protocols:	Report on COVID-19. The Permanent Taskforce's approach has created a framework for platforms, NGOs and other ecosystem stakeholders to effectively work together.
(i) Awareness-raising measures for recipients of the services:	<p>Using COVID-19 as an example, TikTok has the following range of measures which it can deploy depending on the nature of the Medical Misinformation:</p> <ul style="list-style-type: none">• In-app notices proactively directing users to verified information when they search for certain key terms;• Safety Centre resource resources which draws on a range of trusted sources;• Supporting vaccine education programmes with in-app awareness programmes;• TikTok's Transparency Centre has dedicated content on Combating misinformation; and• TikTok's Help Centre contains 'how to' explanations to allow users to learn about its content moderation practices and how to report violative content.
(j) Targeted measures to protect the rights of the child:	Not applicable.
(k) Measures to identify and address inauthentic content and behaviours:	Please see the corresponding section in the Hate Speech section of this Report for details.

12. RISKS TO PUBLIC SECURITY FROM HARMFUL MISINFORMATION/CONTENT

Description of the risk:

- 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) (together “**Public Security Risks**”).
- 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, etc.) or services (e.g., banks, cash machines, etc) are no longer available in a particular location, causing hoarding;
 - 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 Art. 6 of the Charter enshrines for everyone the right to liberty and security of person, and Art. 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.

Key mitigation measures put in place:

- Risk Mitigations - Table 8 sets out a summary of risk mitigation measures in place with reference to Art. 35 (1) (a) to (k) of the DSA.
- Key mitigations for this risk are: (1) TikTok’s policies and proactive content moderation enforcement systems and processes; (2) TikTok’s 24/7 incident management processes to ensure the swift removal of content related to Public Security Risks; and (3) TikTok’s Fact-Checking Programme, with [IFCN-certified](#) independent fact-checking partners across

Europe who assist TikTok in verifying, labelling or removing content.

- As Public Security Risks can be, but are not necessarily localised, TikTok's Fact-checking programme makes sure that any differences are understood and factored into planning and execution of risk mitigation strategies.

Key data relied on:

- TikTok does not report on content detected and removed from the Platform due to Public Security Risks and content removed for these reasons is contained within its reporting on harmful misinformation/content.
- TikTok reported in its Q1 2023 Community Guidelines Enforcement Report that it detected and removed 908,927 pieces of video content globally under TikTok's policies and processes for integrity and authenticity (which comprises harmful misinformation/content as well as spam and fake engagement). In addition:
 - 94.8% of such content was detected and removed by TikTok;
 - 72.8% was removed before there were any views of that content; and
 - 76.6% was removed within 24 hours of upload.
- TikTok's reporting under the [CoPD](#) contains relevant information and metrics on its efforts to combat disinformation:
 - In Q1 & Q2 2023, fewer than 2 in 10,000 views occurred on content identified and removed for violating its policies on harmful misinformation/content (all, not just content related to Public Security Risks): i.e., across the EU, 140,635 videos were removed, with the number of views of such videos in the EU being 1,012,020,899 (the equivalent numbers for the EEA are 142,711 and 1,019,752,855 respectively).³⁰
 - The number of ads removed for violating TikTok's ads policies under the political content ad policy was 390 in the EU/395 in the EEA (for the 6-month period of 1 January to 30 June 2023, as indicated in its CoPD Report).³¹
 - In its Q1 and Q2 CoPD 2023 report TikTok reported on covert influence operations it detected which were attempting to artificially amplify specific viewpoints in the context of the war in Ukraine. TikTok removed a total of 8,358 videos for violating its harmful misinformation/content policy as it relates to Public Security Risks.

Severity:

- Harmful misinformation/content in the context of a real-world crisis event or emergency may cause confusion, panic and could contribute to real-world harm, such as violence, property damage or looting.
- Risks to public security from harmful misinformation/content can be highly localised due to the scale of the underlying events or circumstances that give rise to such risks, but could have a country-wide reach.
- TikTok's trend analysis indicates that risks to public security from harmful misinformation/content typically spike for a short but intense time period before receding.
- Such risks, notably their offline manifestations, can generally be remediated by implementing effective measures to proactively detect and remove content related to Public Security Risks.
- TikTok assesses the severity of Public Security Risks to be moderate, due to: (1) harms and their duration will depend on the nature and scale on the precise facts of real world events, over which TikTok does not have control; and (2) TikTok's proactive efforts to enforce its policies, as demonstrated by the mitigation measures referred to above and in particular to TikTok's 94.8% proactive detection rate. In particular this assessment results from the measures put in place to

³⁰ For the period 1 January - 30 June 2023, as reflected in TikTok's Code of Practice on Disinformation, p. 135 - 137.

³¹ For the period 1 January - 30 June 2023, as reflected in TikTok's Code of Practice on Disinformation, p.4 - 6.

reduce its negative impact of Public Security Risks content on the Platform.

- A challenge in combating harmful misinformation/content is that scenarios giving rise to Public Security Risks tend to evolve quickly and without much advance warning, which places an emphasis on TikTok's capacity for agile responsiveness to quickly detect, contain and mitigate the severity of such evolving risks through its incident management processes.

Probability:

- The circumstances giving rise to public security risks are unpredictable and require constant monitoring of unfolding events to ensure that emerging risks are quickly detected, contained and mitigated. Please see the Case Study in the next section for how TikTok managed recent civil unrest in France.
- TikTok's Ad Policies are designed to protect users from fake, fraudulent, or misleading content. Advertiser accounts and ad content must comply with TikTok's Community Guidelines, which include relevant prohibitions on harmful misinformation/content. TikTok's pre-moderation of ads means any advertising on the Platform is unlikely to include forms of harmful misinformation/content.
- TikTok strives to prevent the upload of, or otherwise remove, content that gives rise to Public Security Risks from the Platform. TikTok assesses that it is possible that there remains a level of Public Security Risks content on the Platform, which is likely to occur at unpredictable, yet specific, times. This is because TikTok implements a range of measures to detect and remove harmful misinformation/content and through in-app interventions that lead users to reliable information sources, such to reduce the probability that such a systemic risk would have a large-scale impact in Europe.

Key stakeholder engagement:

- As a signatory to the original CoPD, TikTok engaged extensively with the CoPD Permanent Taskforce, various working groups together with the European Commission, other platforms, and key stakeholders (including European Digital Media Observatory). As part of its commitments to the CoPD, TikTok will publish a transparency report every six months to provide granular data for European countries about its efforts to combat online misinformation (available [here](#)).
- As part of its Fact-Checking Programme, TikTok works closely with its [IFCN-certified](#) independent fact-checking partners (TikTok's European fact-checkers include Agence France-Presse (AFP), Deutsche Presse-Agentur (DPA), Facta, Logically, Lead Stories, Newtral, Science Feedback, Teyit, and Reuters).
- TikTok has a number of regional [Safety Advisory Councils](#), including for Europe, which are an important source of expert advice and bring outside perspectives to its safety work.

Prioritisation:

- TikTok considers the dissemination of Public Security Risks Content to be a **Tier 2** priority.
- This is due to the unpredictable and potentially isolated circumstances giving rise to such risks in Europe. TikTok acknowledges that such risks have a capacity to quickly evolve into acute risks in the context of an unfolding crisis event or emergency.
- Accordingly, TikTok will continue to enforce its Community Guidelines policies on [Misinformation](#) and [Violent Behaviours and Criminal Activities](#) as normal, and its specialist teams will also remain vigilant and undertake continuous risk monitoring and ensure 24/7 readiness through TikTok's incident management processes, so that it can quickly detect, mitigate and contain emerging such risks as they materialise.

Key further mitigation effectiveness improvements in line with Art. 35 of the DSA:

- **Detection improvements (Art. 35(1)(f)):** TikTok will further expand its Fact-Checking Programme by on-boarding new European-based fact-checking partners and increasing its operational coverage in Europe; and
- **Continued risk monitoring and vigilance (Art. 35(1)(f)):** TikTok will continue to devote cross-functional resources on a priority basis to its processes for handling Medical Misinformation risks.

Risk Mitigations - Table 8	Risks to public security from harmful misinformation/content
TikTok's risk-mitigation measures in accordance with Art. 35(1) of the DSA (a) to (k), plus any other relevant measures:	
(a) Adaptation of feature or platform design, including online interfaces (as defined in Art. 3(m) DSA):	Please see the corresponding section in the Medical Misinformation section of this Report for details.
(b) Adaption of terms and conditions (as defined in Art. 3(u) DSA) and their enforcement:	<p>TikTok prohibits harmful misinformation/content including content related to Public Security Risks on the Platform through a combination of TikTok's Terms of Service and the Community Guidelines, and takes a range of measures to generate awareness and inform users.</p> <p>Under the heading of "Misinformation", TikTok's Community Guidelines prohibit harmful misinformation/content, stating that <i>'We do not allow inaccurate, misleading, or false content that may cause significant harm to individuals or society, regardless of intent'</i>. The Community Guidelines also make it clear that <i>'misinformation that poses a risk to public safety or may induce panic about a crisis event or emergency'</i> is not allowed on the Platform.</p> <p>The Guidelines provide further detail to users about what each of these terms mean and make clear that statements of personal opinion (as long as it does not include harmful misinformation) are permitted.</p> <p>Under the heading of "Synthetic and Manipulated Media" TikTok's Community Guidelines also requires <i>'synthetic media or manipulated media that shows realistic scenes must be clearly disclosed'</i> and makes clear that <i>'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'</i> is not allowed on the Platform.</p> <p>In addition, under the heading of "Safety and Civility", TikTok's Community Guidelines prohibit Violent Behaviours and Criminal Activities, stating that <i>'We do not allow any violent threats, incitement to violence, or promotion of criminal activities that may harm people, animals, or property'</i>.</p>

(c) Adoption of content moderation (as defined in Art. 3(t) DSA) processes:	Please see the 'Key Information about TikTok' section for a description of TikTok's content moderation processes.
(d) Testing and adaption of algorithmic systems, including recommender systems (as defined in Art. 3(s) DSA):	Please see the corresponding section in the Medical Misinformation section of this Report for details.
(e) Adaptation of advertising systems:	<p>Please see the 'Key Information about TikTok' section for a description of TikTok's content moderation processes.</p> <p>In addition, with specific applicability to harmful misinformation/content:</p> <ul style="list-style-type: none"> • TikTok's dangerous misinformation policy expressly prohibits content encouraging people to destroy public property due to catastrophe. For example, content falsely claiming basic necessities (such as food or water) or services (including banks, access to cash machines) are no longer available in a particular location, causing hoarding or over-purchasing; and • TikTok's Dangerous Conspiracy Theory Policy prohibits 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. For example, conspiratorial content that is accompanied by a violent call to action is prohibited on the Platform.
(f) Reinforcing risk detection measures:	<p>Please see the corresponding section in the Election Misinformation section of this Report for details of TikTok's Fact-checking Programme.</p> <p>In addition:</p> <ul style="list-style-type: none"> • TikTok's Trust & Safety Risk Analysis team comprises a designated Risk Analysis team staffed by experienced professionals with backgrounds in cyber intelligence and risk detection. This team's activities include monitoring open-source resources and reporting to cross-functional colleagues on potential new and emerging Public Security Risks; and • TikTok's dedicated Law Enforcement Outreach team engages in outreach with national law enforcement authorities across the EU and with agencies such as Europol. This outreach supports TikTok's risk detection and intelligence gathering on potential new and emerging Public Security Risks.
(g) Cooperation with trusted flaggers:	Please see the corresponding section in the Election Misinformation section of this Report for details.
(h) Cooperation with other platforms through the codes of conduct/crisis protocols:	<p>Please see the corresponding section in the Election Misinformation section of this Report for details.</p> <p>In addition TikTok's CoPD reports include information about Covert Influence Operations it identifies and removes from the Platform.</p>

(i) Awareness-raising measures for recipients of the services:	<p>TikTok has in place the following measures to combat the spread of Harmful Misinformation:</p> <ul style="list-style-type: none">• Safety Centre resource resources which draws on a range of trusted sources;• TikTok's Transparency Centre has dedicated content on Combating misinformation and Countering influence operations;• TikTok's Help Centre provides users with accessible 'how to' explanations of the Platform's user experience, to allow those users to learn about the Platform and troubleshoot issues; and• TikTok has implemented campaigns to address media literacy and combat misinformation about major world events that have significant impact on public security, including armed conflict, natural and man-made disasters.
(j) Targeted measures to protect the rights of the child:	Not applicable.
(k) Measures to identify and address inauthentic content and behaviours:	Please see the corresponding section in the Election Misinformation section of this Report for details.

Case Study: Handling civil unrest in France

Context: In the aftermath of the fatal shooting of a 17-year old Nahel Merzouk (“**Nahel**”) on Tuesday 27 June 2023, there followed a period of civil unrest in France, involving violent protests, property damage and looting. Media reports indicate that Nahel was shot as he drove away from a French police officer. In the days that followed, footage allegedly showing the incident and the subsequent riots were widely shared online, as well as content depicting violent interactions with the police, looting and the use of AI-generated deep fakes of Nahel and the police officer that shot him.

Risk monitoring and readiness: TikTok’s teams became aware of the incident from breaking local media reports and monitoring of other online platforms, as well as through engagement by its Law Enforcement Outreach team with French law enforcement. Led by the Incident Management team, a cross-functional, multidisciplinary team was quickly convened to closely assess the evolving situation.

Measures taken: TikTok took a range of actions to manage and contain this evolving risk over a number of days:

- TikTok’s Incident Management team provided 24/7 coverage, including by conducting sweeps for duplicates of violating video content.
- Policy guidance was issued to the Incident Management team (within 1 hour of the request) to help clarify the types of content that were violating TikTok’s Community Guidelines.
- TikTok’s Law Enforcement Outreach team maintained engagement with local and national French law enforcement throughout this evolving situation.
- TikTok’s Risk Analysis team assessed trending content, [REDACTED] on other platforms to assess potential violative content trends and detect emerging risks, including by identifying [REDACTED] associated with violating content.
- Newly created accounts were assessed where they appeared dedicated to the promotion of illegal activity, such as violent riots, destruction of property, and shoplifting, notably through [REDACTED].
- TikTok handled an increased volume of user reports for Community Guidelines violations (mainly through its French and Arabic language moderation teams).
- For content that had previously been detected as content which violates TikTok’s Community Guidelines, deduplication technologies were used which enabled TikTok to recognize copies or near copies of such content, and prevent re-posting.
- TikTok’s approach to removing harmful content has been balanced with protection of freedom of expression and recognition that not all content about this evolving situation was violative. For example, this involved content calling for justice or peaceful protests or media coverage explaining the events, which may involve depictions of protest, riots and property destruction.
- TikTok received a flash report on misinformation trends around France protests from one of its fact-checking partners (Agence France Presse).

Summary of action taken: Based on the above measures, the following moderation action resulted in:

- Over 3,000 videos being removed from the Platform;
- Over 900 videos being labelled with a mask layer and ineligible for recommendation;
- Multiple accounts being banned (including for banned for the illegal trade of firearms (in this case, often explosives such as mortars or fireworks used for the riots).

13. RISKS TO FUNDAMENTAL RIGHTS

Description of the risk:

- 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 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.
- 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).

Key mitigation measures put in place:

- Risk Mitigations - Table 9 sets out a summary of risk mitigation measures in place with reference to Art. 35(1)(a) to (k) of the DSA.
- Key mitigations for this risk are: (1) the checks and balances on TikTok's policies and proactive content moderation systems and process which includes the public interest exceptions which apply; and (2) the role of the Platform Fairness team and its outreach and learning with external partners and experts. TikTok also notes that the cumulative effect of all other general and specific mitigations referenced in this Report as also supplementing the protection of Fundamental Rights.
- TikTok carefully considers situations where a difference in regional/local law appears to conflict with international human rights standards and seeks solutions to uphold its users' rights to free expression and privacy.

Key data relied on:

- TikTok's mission is to inspire creativity and bring joy by enabling creative expression. TikTok's Trust & Safety teams dedicate significant resources to ensure that it offers users an environment that embraces freedom of expression and protects the fundamental rights of TikTok's community. This is reflected in TikTok's quarterly [Community Guidelines Enforcement Reports](#). In Q1 2023 total videos removed represent about 0.6% of all videos uploaded to TikTok.
- Additionally, over a 12-month period from April 2022 to March 2023, TikTok's teams and technology have proactively removed 96.3% of identified violative video content before it was reported to TikTok; removed 91.9% of violative video content within 24-hours of it being posted; and removed 86.4% of violative videos before they received any views. TikTok is proud of these numbers, but constantly strive to do more to ensure a positive experience.

Severity:

- TikTok is conscious that Fundamental Rights are varied and careful balancing is required when those rights are exercised in practice on the Platform. For example, the right to

freedom of expression, may infringe on another user's right to non-discrimination. Unfettered expression could also result in a negative impact on the freedom of expression by marginalised groups (e.g., where the dominant voices in society monopolise opportunities for expression and suppress other views).

- Notwithstanding the central role held by the freedom of expression in the international legal framework, this “*freedom is not absolute*” and “*can be lawfully restricted in order to balance it against other fundamental rights*”.³²
- These considerations therefore require careful balance and thoughtful mitigations. As these rights persist for all users of the Platform, regardless of the duration of use, TikTok therefore considers these rights in the application of its decisions about the content of its Community Guidelines and their enforcement.
- The duration and remediability of the impacts of the dissemination of content that undermines users' Fundamental Rights will depend on the specific nature of the content and the operation of content moderation policies and processes.
- TikTok assesses the potential severity of risks to the six fundamental rights outlined above to range from very low to moderate. This is due to the variability of harm, duration, and remediability across categories of Fundamental Rights, especially taking into consideration that protection of certain rights may infringe on others.

Probability:

- Given the nuanced and competing factors above, it is challenging to assess the probability of negative effects on users' exercise of their Fundamental Rights on the Platform. Given the reach of TikTok and the wide variety of content available on the Platform, it would be possible that negative effects could occur were TikTok to fail to protect and balance each of the Fundamental Rights described above.
- Consistent with its mission “to inspire creativity and bring joy”, TikTok gives special consideration to the impact and potential negative effects on freedom of expression. TikTok employs various measures with a view to striking the appropriate balance between tackling egregious illegal content and supporting the fundamental right of free expression. To do this TikTok implements internationally recognised human rights due diligence frameworks to make sure that measures taken are necessary and proportionate. For example, under TikTok's Community Guidelines, TikTok allows for policy exceptions in the context of certain harmful or illegal content, such as survivors discussing their own experiences with youth exploitation and abuse, or educational and documentary content that raises awareness of the harms caused by violent and hateful actors.
- TikTok does not publish reports into content removals for breaches of Fundamental Rights. Instead, TikTok publishes quarterly [Community Guidelines Enforcement Reports](#) report on the volume of content removed for violating all its Community Guidelines. The Community Guidelines are TikTok's expression of what content is/not permissible and as such is the product of TikTok's consideration of the risks described above. On a six-monthly basis, TikTok also publishes [Government Removal Requests Report](#); and [Information Requests Report](#).
- TikTok's Ad Policies are designed to protect users from fake, fraudulent, or misleading content. A consideration of fundamental rights is enshrined in TikTok's Ad Policies, which are rooted in law. TikTok's Community Guidelines also expressly address the competing nature of those interests and the need to balance competing interests in its enforcement actions. Therefore, in its moderation of ads, TikTok has regard to individuals' Fundamental Rights.
- TikTok considers that it is possible that there remains a level of risk to users' Fundamental

³² European Law Institute, [Freedom of Expression as a Common Constitutional Tradition in Europe](#) (August 2022).

Rights. These risks can be mitigated through ensuring that TikTok's operation of its content moderation policies and procedures properly consider whether there is any unjustifiable impact on any relevant Fundamental Right.

Key stakeholder engagement:

- TikTok's six regional [Safety Advisory Councils](#), including for Europe, are an important source of expert advice and bring outside perspectives to TikTok's safety work.
- TikTok's ongoing memberships with industry groups allows it to engage with peers on current and upcoming trends. TikTok also participates in the Business for Social Responsibility's Human Rights Working Group and is engaging with the United Nations Human Rights Office of the High Commissioner on its B-Tech Project, which provides guidance and resources for implementing the United Nations Guiding Principles on Business and Human Rights in the technology space.
- TikTok also operates a Fairness and Inclusion Advisory Network: This is a small group of diverse external experts (some of whom are also part of the Safety Advisory Councils) that advise the Platform Fairness team on relevant projects and initiatives.
- TikTok regularly engages with NGOs to develop policies and in app features as members of its Community Partner Channel. Examples include WITNESS, E-enfance, Amedeu Antonio Stiftung, ZARA Zivilcourage, Tito de Morais, HackerOne, Cyberbullying Research Center, and Internet Matters.

Prioritisation:

- TikTok considers fundamental rights to be a **Tier 3** priority.
- This is due to the potential individual and societal harms and the range of measures already implemented to ensure the exercise and protection of fundamental rights (and noting that some key rights, such as the rights of the child, are more specifically considered under several other risk categories within this Report).
- TikTok will continue to implement targeted measures to ensure the Platform continues to be a space where its users are empowered to exercise their fundamental rights, led in particular by specialist Product, Trust & Safety and Privacy teams.

Key further mitigation effectiveness improvements in line with Art. 35 of the DSA:

- **Detection developments (Art. 35(1)(f)):** TikTok aims to expand its collaboration with external parties to enhance intelligence-gathering capacity, including detection of emerging risks.
- **Media literacy (Art. 35(1)(i)):** TikTok aims to supplement its media literacy measures to generate awareness of Fundamental Rights issues and to encourage users to report content undermining Fundamental Rights.
- **Continued risk monitoring and vigilance (Art. 35(1)(f)):** TikTok aims to devote further resources to scenario-planning and readiness to address risks to Fundamental Rights, including those referenced in the Hate Speech module. This should include developing an approach to ongoing human rights due diligence in line with UN Guiding Principles.

Risk Mitigations - Table 9	Risks to fundamental rights
TikTok's risk-mitigation measures in accordance with Art. 35(1) of the DSA (a) to (k), plus any other relevant measures:	
(a) Adaptation of feature or platform design, including online interfaces (as defined in Art. 3(m) DSA):	Please see the corresponding section in the Hate Speech section of this report for details.
(b) Adaption of terms and conditions(as defined in Art. 3(u) DSA) and their enforcement:	<p>TikTok has set out a <u>Commitment to Human Rights</u>, which is informed by several international human rights frameworks. This Commitment serves as a touchpoint and guide for TikTok's Terms of Services, content moderation practices such as its Community Guidelines and how they are enforced and localised, user data and privacy policies, policies related to its supply chain and business partnerships, and its transparency practices.</p> <p>TikTok prohibits content undermining the Fundamental Rights on the Platform through a combination of its <u>Terms of Service</u> and <u>Community Guidelines</u>, and takes a range of measures to generate awareness and inform users.</p> <p>In addition to the extracts from the Community Guidelines referred to in other sections of this Report, they also set out TikTok's position on Fundamental Rights in its Community Principles, under the themes of "Balance", "Dignity", and "Fairness". The Community Principles make clear that TikTok's <i>"content moderation principles and practices are informed by the UN Guiding Principles on Business and Human Rights and the Santa Clara Principles, and [that we] we seek to align with international legal frameworks, such as the International Bill of Human Rights and the Convention on the Rights of Children."</i> Importantly, the Community Principles recognize that <i>"sometimes these principles may be in tension with each other, and we make trade-offs carefully."</i></p> <p>TikTok's terms and conditions also include its 'Virtual Items Policy' which covers their purchase and in-app gifting by users.</p>

<p>(c) Adoption of content moderation (as defined in Art. 3(t) DSA) processes:</p>	<p>Please see the 'Key Information about TikTok' section for a description of TikTok's content moderation processes.</p> <p>TikTok (as described in its Community Guidelines) applies a Public Interest Exception. This is because content that would otherwise violate its rules could be in the public interest to view. This does not simply refer to what the public may be interested in. Public interest refers to topics that inform, inspire, or educate the community and enhance deliberation about matters of broad collective significance. TikTok may allow content to remain on the platform under one of the following public interest exceptions: Documentary, Educational, Medical and Scientific, Counterspeech, Satirical or Artistic. TikTok may add extra safety measures to some content allowed under a public interest exception.</p>
<p>(d) Testing and adaption of algorithmic systems, including recommender systems (as defined in Art. 3(s) DSA):</p>	<p>Please see the corresponding section in the Hate Speech section of this report for details.</p>
<p>(e) Adaptation of advertising systems:</p>	<p>Please see the 'Key Information about TikTok' section for a description of TikTok's content moderation processes.</p> <p>In addition:</p> <ul style="list-style-type: none"> • With reference to consumer protections, TikTok's Ads Policies require that all ads on the Platform are legal and not misleading. TikTok's ads moderation processes are similar to those set out in relation to user generated content; • TikTok's Anti-Discrimination Ad Policy prohibits advertisers from using ad products to discriminate against people wrongfully; and • TikTok's Privacy Policy contains information about the data that it collects for advertising purposes.
<p>(f) Reinforcing risk detection measures:</p>	<p>As well as all the teams otherwise referred to in this Report, TikTok operates a number of specialist teams that deal with Fundamental Rights risks in particular and who play a role in reinforcing risk detection measures:</p> <ul style="list-style-type: none"> • The Platform Fairness team, which sits within Trust & Safety, is committed to the equitable treatment of TikTok's stakeholders across product policies, the moderation process, and in-app features. It focuses on promoting baseline human rights, identifying gaps and biases in policy language and enforcement, and working to ensure new product features are built to mitigate against biases. TikTok has implemented a compliance review process for new AI models, which incorporates guidance on unfair bias prevention from its Platform Fairness team; and • The Privacy and Responsibility Product Management team focuses on issues relevant to minor protection; data transparency, management and access; and user choices over personalisation, interaction and discoverability.

(g) Cooperation with trusted flaggers:	TikTok operates a trusted flagger programme, the Community Partner Channel, whereby it has established separated reporting channels for various credible and trusted NGOs, charities and regulatory authorities to report harmful content to us. These partners report on a variety of content, including that which could undermine Fundamental Rights, such as discriminatory or dehumanising speech, and content that could confuse or manipulate users. These include organisations who focus on reducing racism, xenophobia, bullying, frauds and scams and promote freedom of expression and protection of personally identifiable information.
(h) Cooperation with other platforms through the codes of conduct/crisis protocols:	Not applicable.
(i) Awareness-raising measures for recipients of the services:	<p>TikTok undertakes the following measures:</p> <ul style="list-style-type: none"> • TikTok's Transparency Centre has dedicated content explaining TikTok's approach to <i>Keeping People Safe</i>, which includes separate articles that explain 'Our approach to content moderation'; • TikTok's Help Centre contains 'how to' explanations to allow users to learn about its content moderation practices and how to report violative content; • TikTok publishes its Privacy policy which details (among other things) how it collects, uses, shares users' information, how it is processed and users' rights and choices; and • TikTok performs a range of digital literacy activities, as described elsewhere in this Report.
(j) Targeted measures to protect the rights of the child:	Please see the corresponding sections in all the in the Hate Speech section of this report for details.
(k) Measures to identify and address inauthentic content and behaviours:	Please see the corresponding section in the Hate Speech section of this report for details.

Deep Dive: Striking a balance between preventing harm and enabling expression

Overview: TikTok's mission is to inspire creativity and bring joy by enabling creative expression. To help ensure a safe, trustworthy, and vibrant experience, TikTok maintains a set of Community Guidelines that include rules and standards for using TikTok. A complex issue for any online platform is striking an appropriate and proportionate balance between preventing harm and enabling freedom of expression. This Case Study explains how TikTok strives to strike this balance, in particular through its Community Guidelines.

TikTok's Community Principles: TikTok's Community Guidelines are the starting point when it comes to how TikTok frames and shapes its content moderation strategies and practices. They are informed by international legal frameworks, industry best practices, and input from community, safety and public health experts, and its [regional Advisory Councils](#).

Under its [Community Guidelines](#), TikTok has eight guiding community principles that help embody its [commitment to human rights](#). TikTok's principles are centered on balancing expression with harm prevention, embracing human dignity, and ensuring its actions are fair. The first two principles in particular seek to strike a balance between:

1. **Preventing harm:** *TikTok's primary focus is keeping its community safe, fostering inclusivity, and ensuring TikTok is a place for joy. TikTok considers the many ways that content or behaviours may cause harm to individuals or its diverse community. This includes physical, psychological, financial, privacy, and societal harms. To strike the right balance with free expression, it restricts content only when necessary and in a way that seeks to minimise the impact on speech.*
2. **Enabling free expression:** *The creativity unlocked by [expression](#) is what powers TikTok's vibrant community. It honours this human right by providing the opportunity to share freely on its platform and by proactively removing harassing behaviour that can inhibit creator speech. However, free expression is not an absolute right – it is always considered in proportion to its potential harm. It also does not extend to a right to have your content amplified in the For You Feed.*

The principles shape TikTok's day-to-day work and guide how it approaches difficult enforcement decisions. This may involve situations where harm prevention and enabling expression pull in different directions and appear to be in conflict. In such cases, TikTok strives to strike a fair and proportionate balance between the rights concerned - for example, to ensure space for free expression, it may allow some more latitude for social critique of public figures.

Public interest exceptions: TikTok's [Community Guidelines](#) recognise that some content that would otherwise violate its rules can be in the public interest to view. The section on [Public Interest Exceptions](#), explains that this does not simply refer to what the public may be interested in. Public interest refers to topics that inform, inspire, or educate the community and enhance deliberation about matters of broad collective significance. TikTok may allow content to remain on the Platform under one of the following public interest exceptions:

- Documentary
- Educational

- Medical and Scientific
- Counterspeech
- Satirical
- Artistic

TikTok's approach to content moderation uses the same criteria, no matter who creates it. The most important factor in looking at public interest exceptions is context, such as captions, voice over, and similar signals. TikTok encourages creators to clearly show the context to help in its review process.

Examples of public interest exceptions: TikTok's Community Guidelines policy on [Hate Speech and Hateful Behaviours](#) makes clear that it does not allow any hateful behaviour, hate speech, or promotion of hateful ideologies. Similarly, the Community Guidelines policy on [Violent and Hateful Organisations and Individuals](#) makes clear that TikTok does not allow the presence of violent and hateful organisations or individuals on the Platform, and that it does not allow anyone to promote or materially support violent or hateful actors. To balance harm prevention with freedom of expression, the Community Guidelines articulate the following public interest exceptions in relation to hate speech and violent extremism:

In the context of [Hate Speech and Hateful Behaviours](#):

- Self-referential slurs used by a member of a group with that particular protected attribute; and
- Educational and documentary content raising awareness against hate speech.

In the context of [Violent and Hateful Organisations and Individuals](#):

- Discussing a violent political organisation (as long as there is no mention of violence); and
- Educational and documentary content that raises awareness of the harms caused by violent and hateful actors.

Local nuance and evolving speech: Assessing speech can be very complex and localised due to the regional, cultural and linguistic context which leads to differences across the EU in terms of how such speech is formulated and expressed. For example, those who may be targeted by hate speech or hateful behaviour, and the manner in which such content is expressed may differ from one country to another. For these reasons, through its regional Trust & Safety teams TikTok adopts a localised and nuanced approach to tackling such risks. This includes having relevant linguistic capacity to understand and moderate content in a wide range of languages. TikTok's regional policy teams play a key role in helping to understand the local and cultural nuances, such as emerging risks including new local slang or slur terms, and facilitating training and awareness of updates. This work facilitates TikTok's content moderation teams to take a more regionally nuanced and informed approach to detection of risks and content moderation more generally.

14. RISKS OF INTELLECTUAL PROPERTY INFRINGING CONTENT

Description of the risk:

- TikTok understands the term “**IP-Infringing Content**” to mean content that is created and disseminated in breach of copyright or other intellectual property 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.

Key mitigation measures put in place:

- Risk Mitigations - Table 10 sets out a summary of risk mitigation measures in place with reference to Art. 35 (1) (a) to (k) of the DSA.
- Key mitigations for this risk are: (1) TikTok uses best efforts to agree licences with rightsholders and/or collective management organisations. By way of example, it has agreed licences from all major music rightsholders worldwide; (2) TikTok operates a notice and takedown process, which covers copyright and trademark forms of IP-infringing Content and covers all major European languages; and (3) 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.

Key Data relied on:

- TikTok’s [Intellectual Property Removal Request Reports](#) reports on notice and takedown reports actioned on IP-Infringing Content (comprising in this case copyright, counterfeit, and trademark violations). The latest global report (for the period July to December 2022) states that TikTok took the following action in relation to items of IP-Infringing Content:
 - For copyright content, a total of 168,141 requests were made, of which 95,479 requests were successful copyright removal requests; and
 - For trademark content, a total of 19,239 requests were made, of which 11,624 successful trademark removal requests.³³

³³ For more information please see TikTok’s [Intellectual Property Removal Requests Report](#).

Severity:

- Infringing intellectual property rights is unlawful across the EU. Unauthorised use or sharing of creative content threatens rightsholders' creative investments, and can cause economic harm.
- Users who view IP-Infringing Content may share it, thus exacerbating the risk of harm to rightsholders.
- If intellectual property rights are infringed at scale, this could chill innovation and negatively impact the marketplace within Europe. Such risks can generally be mitigated through licensing arrangements, notice-and-takedown mechanisms and collaborative technical solutions.
- This risk is also largely remediable, in particular compared to other risks within the illegal content category, where harms are psychological or physical.
- TikTok's key data above demonstrates that TikTok's mitigation measures allow it to respond to reporting of IP-Infringing Content.
- TikTok assesses the potential severity of the risk of dissemination of IP Infringing Content on the Platform to be moderate due to: (1) the potential for economic loss for creators and authors as well as the wider European marketplace; and (2) the mitigation measures described above which are put in place to restrict the scale of the dissemination of IP-Infringing Content.

Probability:

- TikTok strives to prevent the dissemination of IP-Infringing Content on the Platform. TikTok considers that it is possible that there will remain some level of IP-Infringing Content on the Platform. This assessment is based on several factors including: the tools available to rightsholders to report IP-Infringing Content, the number of notices received from rightsholders and actioned by TikTok in comparison with the overall volume of content on the Platform.
- TikTok's Ad Policies are designed to protect users from fake, fraudulent, or misleading content. TikTok's Ad Policies prohibit ads that violate or infringe upon the rights of any third party. TikTok's pre-moderation of ads means any advertising on the Platform is unlikely to include IP-Infringing Content.

Key stakeholder engagement:

- TikTok engages with various industry bodies and external experts who provide information and insights that simultaneously inform and confirm the extent of the IP-Infringing Content risk, and assist with the development of TikTok policies. Examples of such engagement include:
 - **Audiovisual Anti-Piracy Alliance (AAPA):** AAPA is an industry group, including rights owners and broadcasters, aiming to tackle audiovisual piracy in Europe and beyond through effective lobbying, supporting law enforcement and building partnerships;
 - **Union des fabricants (UNifab):** UNifab is a French association that promotes and defends intellectual property; and
 - **Danish Rights Alliance (DRA):** DRA is an interest organisation representing more than 100,000 Danish rightsholders in the film, music, literature, image, design and media space.

Prioritisation:

- TikTok considers risks of dissemination of IP-Infringing Content to be a **Tier 3** priority.
- This is due to the largely remediable economic harms involved in such illegal content. Further, the risk is not highly dynamic in nature and there are established on and off Platform mitigation measures in place.

- TikTok will continue to enforce its prohibition on such content, in particular through its notice-and-action mechanisms, and to actively engage with rightsholders to protect their rights.

Key further mitigation effectiveness improvements in line with Art. 35 of the DSA:

- **Detection developments (Art. 35(1)(f)):** TikTok will continue to further develop its internal tools to detect and prevent copyright infringing activities where rightsholders have provided TikTok with the relevant and necessary information.
- **External engagement:** TikTok will continue to use best efforts to agree licences with rightsholders under Art. 17 of the Copyright Directive.
- **Continued risk monitoring and vigilance (Art. 35(1)(f)):** TikTok will continue to devote cross-functional resources on a priority basis to its processes for handling IP-Infringing Content risks. In addition, TikTok shall continue to collect and monitor relevant data as part of its transparency reporting obligations under the DSA.

Risk Mitigations - Table 10	Risks of intellectual property infringing content
TikTok's risk-mitigation measures in accordance with Art. 35(1) of the DSA (a) to (k), plus any other relevant measures:	
(a) Adaptation of feature or platform design, including online interfaces (as defined in Art. 3(m) DSA):	<p>TikTok implements a range of measures to prevent and mitigate the risk of the dissemination IP-Infringing Content on or through the Platform, including through the following features, Platform design, and relevant technical safety measures:</p> <ul style="list-style-type: none"> • TikTok provides various means for rightsholders and representatives to request notice and takedown, blocking or licensing through webforms and other copyright tools; and • Certain rightsholders may also use 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.
(b) Adaption of terms and conditions (as defined in Art. 3(u) DSA) and their enforcement:	<p>TikTok prohibits unauthorised IP-Infringing Content on the Platform. This is done through a combination of TikTok's Terms of Service, Community Guidelines, and Intellectual Property Policy and takes a range of measures to generate awareness and inform users.</p> <p>Under the heading of "Intellectual Property", TikTok's Community Guidelines state that <i>'We do not allow posting, sharing, or sending any content that violates or infringes someone else's copyrights, trademarks or other intellectual property rights. We may remove infringing user content.'</i> The Guidelines provide further detail to users about what each of these terms mean.</p>

	<p>The Intellectual Property Policy (and a corresponding policy for ads) contains a comprehensive overview of how TikTok defines copyright and trademark and protects those rights. It provides information for rightsholders and users on infringement notifications/counter-notifications, and clarifies the exceptions under EU law that may allow users to use copyright protected works without authorisation. It also sets out the penalties for intellectual property violations on the Platform, which include removal of content, and suspension or termination of accounts.</p>
<p>(c) Adoption of content moderation (as defined in Art. 3(t) DSA) processes:</p>	<p>There is no central database of copyright works in Europe. TikTok therefore relies on rightsholders taking action to identify their works to it and has a process to enable rightsholders to do so. In this way, TikTok offers rightsholders a specific notice and takedown process for IP-Infringing Content, which operates as follows:</p> <ul style="list-style-type: none"> • Rightsholders can notify TikTok of an alleged violation of the works or registered trademarks or copyrights. TikTok assesses the request and removes infringing content where a violation is deemed to have occurred; • The Global IP Operations team at TikTok then performs an assessment of whether infringement has occurred, which involves a careful analysis of relevant legal tests and potential exceptions; • If content reported is IP-Infringing Content then the account may be immediately terminated (which may occur if the account appears to have been set up with the sole purpose of infringing protected works). Alternatively, the account may be issued a warning under TikTok's repeat infringer policy; • TikTok notifies users where content is removed and the user is given the the opportunity to file a counter-notification to appeal the decision, which is provided within the TikTok app and through a webform; and • Content that is considered to be IP-Infringing Content and which has not been successfully appealed are ingested into a seed bank if they meet certain criteria. If a new video exactly matches with the seed bank video, the new video will be prevented from being made available in the EU (in line with Art. 17(4)(c) of the Copyright Directive). <p>The process for reporting and taking down IP-Infringing Content in advertising is similar to the process outlined for user generated content above. Additional measures apply for responding to reports about counterfeit goods in livestreams.</p>
<p>(d) Testing and adaption of algorithmic systems, including recommender systems (as defined in Art. 3(s) DSA):</p>	<p>TikTok's personalised FYF is one of the primary means through which users consume video content on the Platform.</p> <p>TikTok removes IP-Infringing Content following a successful notice and take down procedure and such content cannot therefore be displayed on the FYF. In addition, reproduced or unoriginal content that is imported or uploaded without any new or creative edits (such as content with someone else's visible watermark or superimposed logo) is ineligible for the FYF, in accordance with TikTok's Community Guidelines.</p>

(e) Adaptation of advertising systems:	<p>Please see the 'Key Information about TikTok' section for a description of TikTok's content moderation processes.</p> <p>In addition, TikTok's Ad Policies prohibit ads that violate or infringe upon the rights of any third party, including trademarks, copyright and other proprietary rights.</p>
(f) Reinforcing risk detection measures:	<p>TikTok operates a number of specialist teams that deal with IP-Infringing Content in particular and who play a role in reinforcing risk detection measures.</p> <p>The Global IP Operations team aims to quickly and effectively process copyright and trademark takedown requests from rightsholders or their authorised representatives, and to remove any infringements related to user generated content from the Platform. The team consists of IP Operations specialists who are responsible for the reactive aspect of TikTok's IP protection efforts including all blocking and licensing requests for TikTok under Art. 17 of the Copyright Directive.</p>
(g) Cooperation with trusted flaggers:	Not applicable (as IP-Infringing Content is reported by or on behalf of rightsholders).
(h) Cooperation with other platforms through the codes of conduct/crisis protocols:	Not applicable.
(i) Awareness-raising measures for recipients of the services:	TikTok undertakes the following measures: TikTok's Help Centre contains 'how to' explanations to allow users to learn about its content moderation practices and how to report violative content. The content also provides a broad overview of trademark and counterfeiting and a guide to copyright on the Platform.
(j) Targeted measures to protect the rights of the child:	Not applicable.
(k) Measures to identify and address inauthentic content and behaviours:	Please see the relevant section in the Hate Speech section of this Report.

ANNEX 1 - RISK ASSESSMENT METHODOLOGY

TikTok designed and implemented a bespoke risk assessment methodology in order to carry out its benchmark systemic risk assessment of whether, how, and to what extent systemic risks may stem from the design or functioning of the Platform, or the use made of it.

This methodology takes into account a broad range of sources, including:

- The text of the DSA, in particular Art. 34 (and the relevant recitals);
- Risk assessment standards, principles and best practises (including the UN Guiding Principles on Business and Human Rights and ISO risk management standards);
- Equivalent or similar regulatory obligations (such as DPIAs under the GDPR);
- Relevant regulatory guidance from other content and safety regulatory regimes; and
- Consultations with a range of relevant internal and external experts.

The key elements of the methodology included:

- The scoping of systemic risks with reference to all matters referred to in Art. 34(1) of the DSA, with further reference to EU law and other sources, in order to define TikTok's risk categories;
- Assessing (from Art. 34(1)(b) of the DSA and the Charter) the fundamental rights that may be relevant to each risk category;
- Designing and documenting a modular workstream and document drafting approach for each category of risk in order to generate the underlying risk assessments (in accordance with Arts. 34 and 35 of the DSA and their relevant recitals), which are now summarised in this Report;
- Assessing the nature of the risk;
- Articulating the range of mitigation measures that TikTok operates, with reference to Art. 35(1) (a) to (k) and the applicability of fundamental rights;
- Assessing the severity (on a scale of very low, low, moderate, high or very high) and probability (on a scale of very unlikely, unlikely, possible, likely or highly likely) of each category of risk;
- Assigning each systemic risk to a Tier in order to articulate and inform TikTok's prioritisation of measures to address the risk ;
- Identifying any reasonable and proportionate mitigation measures that will be implemented in order to further reduce risk; and
- Identifying those bodies with whom TikTok consults in relation to each category of risk.

Governance and oversight of the risk assessment process was established through:

- A review group made up of senior team leads with relevant experience and subject matter expertise across various of the risk categories;
- An oversight group to provide executive oversight, leadership and strategic direction in relation to TikTok's compliance with content / safety regulatory obligations in Europe;
- The board of directors of TikTok Ireland, which approved the findings of the risk assessments; and
- TikTok's DSA compliance function.

ANNEX 2 - HOW TO USE THIS REPORT

The geographic scope of this Report:

- This Report has been prepared in compliance with Arts. 34, 35 and 42 of the 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 Arts. 34, 35 and 42 of the 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 TikTok's DSA Day 1 on 28 August 2023 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-2023-1/>).

* * *