Code of Practice on Disinformation – Report of TikTok for the period 16 June - 16 December 2022

Table of Content

Executive summary

About TikTok

TikTok's mission is to inspire creativity and bring joy. Millions of people around the world come to TikTok to create, share, discover and connect, and we're committed to maintaining a safe, inclusive and authentic environment for our growing community.

We treat disinformation with the utmost seriousness and are committed to preventing its spread, while elevating authoritative information and investing in media literacy to help get ahead of the problem at scale. At TikTok we place considerable emphasis on proactive content moderation and so are proud that the vast majority of violative content is identified and removed proactively before it receives any views or is reported to us. Our own transparency reports reveal that violating video removals make up about 1% of total videos on TikTok. Violations of our harmful misinformation policies make up less than 1% of overall video removals within the European Economic Area (EEA). However, we are committed to continuing to keep pace with evolving issues that affect our users.

TikTok is currently working to implement the Digital Services Act (DSA), under which the Code will find a new legislative home. We are fully committed to implementing the DSA's provisions. We look forward to continuing to work with the European Commission and other stakeholders to provide greater transparency into the work that we do to keep our users safe across a range of key areas, including disinformation. The publication of this report is a critical step forward into a new era of platform regulation. TikTok is fully committed to embracing this change and meeting its obligations.

The Code of Practice on Disinformation- baseline report

TikTok was a signatory of the original Code and has embraced the strengthened Code from the negotiation phase through to implementation. TikTok has been a proactive participant in the Permanent Taskforce and subgroups set up under the Code. TikTok is proud to publish this first report with comprehensive data from the period of Q4 2022 (unless otherwise stated). We also welcome the opportunity to highlight our ongoing work to combat harmful disinformation about the Covid-19 pandemic and the war in Ukraine.

The revised Code of Practice on Disinformation represents an innovative approach to tackling the spread of disinformation. The Permanent Taskforce's collaborative and highly informed approach has created a framework for platforms, NGOs and other ecosystem stakeholders to effectively work together, recognising that there is unlikely to ever be a finish line. We share the European Commission's hope that this approach will materially reduce harmful content now and in the future.

While we are proud of our efforts for users in the EU and EEA (both prior to the adoption of the Code, and as part of our implementation of it), we acknowledge that we still have more work to do. Our commitment is that we will remain a proactive and engaged signatory and that we will continue to work to meet the commitments that we have signed up to.

Our policies

Our Integrity and Authenticity policies aim to promote a trustworthy, authentic experience for our users. Our policies are framed around harmful misinformation (rather than making a distinction between disinformation and misinformation) and deceptive behaviours.   Our harmful misinformation policies prohibit content that may mislead our community about civic processes, public health, or safety. These policies can be applied to a wide range of content, and that's by design; this content is constantly changing, often based on what's happening in the world.

TikTok provides transparency about harmful misinformation content removals through our Community Guidelines Enforcement Reports, which we have been publishing since 2020. In addition to removing content that is inaccurate and has the potential to harm our users or community, we also tackle deceptive behaviour by removing accounts that seek to mislead people or attempt to use TikTok to deceptively sway public opinion. These activities range from inauthentic or fake account creation, to more sophisticated efforts to undermine public trust. These actors never stop evolving their tactics, and we continually strengthen our policies as we detect new types of content and behaviours.

Enforcing our policies

At TikTok, we have tens of thousands of safety professionals who work together to enforce our Community Guidelines. To do this effectively at scale, we continue to invest in our automated review process as well as in people and training. We rely on automated moderation when our systems have a high degree of confidence that content is violative so that we can swiftly remove violations of our policies.

Disinformation differs from other content issues. Context and fact-checking are critical to consistently and accurately enforcing our misinformation policies. Therefore, while we use machine learning models to help detect potential misinformation, ultimately our approach today is having our moderation team assess, confirm, and remove misinformation violations. We have specialised misinformation moderators who have enhanced training, expertise, and tools to take action on misinformation. This includes direct access to our fact-checking partners who help assess the accuracy of content.

Demonetisation

As TikTok is only a few years old, creator monetisation opportunities are at a relatively early stage of maturity by industry standards. We are fully committed to rewarding creativity and equipping creators with the best tools to find their community, grow their careers and unlock exciting opportunities both on and off the platform. Like all users of our platform, participants in content monetisation programs must adhere to our Community Guidelines, including our Integrity and Authenticity policies. Those policies make clear that we do not allow activities that may undermine the integrity of our platform or the authenticity of our users. They also make clear that we remove content or accounts, including those of creators, that involve misleading information that causes significant harm or deceptive behaviours. In certain scenarios, we may remove a creator's access to a creator monetisation feature.

Advertising content must also comply with our advertising policies. We have measures in place to detect and remove content that violates our policies. We are a GARM (the Global Alliance for Responsible Media) member, and together with our fellow members and signatories, are committed to removing harmful disinformation from monetisation as part of our commitment to upholding the GARM Framework.

Transparency of Political Advertising

We do not allow political actors to place advertising on TikTok and nor do we allow adverts or landing pages which "reference, promote or oppose a candidate for public office, current or former political leader, political party, or political organisation. [adverts] must not contain content that advocates a stance (for or against) on a local, state, or federal issue of public importance in order to influence a political outcome." We allow cause based advertising and public services advertising from government agencies, non-profits and other entities if they are not driven by partisan political motives. We will continue to monitor the evolution of the draft EU Regulation on the Transparency and Targeting of Political Advertising and will carefully consider any further changes to our policies this might require, and consequently any reporting under this Code.

Ensuring the Integrity of Services

Our Integrity and Authenticity policies robustly prohibit deceptive behaviours such as fake accounts and deceptive deep fake videos and we have a range of tactics, techniques and procedures to enforce those policies.

In addition, we have evolved our approach to detecting and taking enforcement action against covert influence networks. TikTok's policies do not allow coordinated attempts to influence or sway public opinion while also misleading individuals, our community, or our systems about an account's identity, approximate location, relationships, popularity, or purpose. We publish our findings on covert influence operations (which started in Q3 2022) through our Community Guidelines Enforcement Reports. We are building and testing processes that can provide this information at the requested frequency with a high degree of fidelity. As a result of this ongoing work, this baseline report comprises data for June to September 2022 only.

Empowering Users

If a user encounters content that violates our Community Guidelines, our in-app reporting tools are accessible and simple to use. Together with systematically removing content from our platform that violates our Integrity and Authenticity policies, we have dedicated significant resources to increasing the number of in-app measures that show users additional context on certain content or redirect them to authoritative information. We currently make those tools available in 21 official languages (plus, for EEA users, Norwegian and, as the spoken language of Liechtenstein, German). We continue to be involved in a variety of on and off-platform media literacy campaigns, often targeted around Member State national elections or ongoing conflicts (for example, the war in Ukraine).

The "For You" feed is the interface users first see when they open TikTok. It's central to the TikTok experience and where most of our users spend their time exploring the platform. We make clear to users that their "For You" feed is based on a personalised recommendation system. This baseline report describes how we build safety into our recommendations. As well as removing harmful misinformation content that violates our Community Guidelines, we take steps to avoid recommending certain categories of content that may not be appropriate for a general audience, including general conspiracy theories. We have also introduced the "Why this video" feature, which empowers our users to better understand why a particular video has been recommended to them.

Empowering Researchers

We recognise the important role of researchers in helping to identify disinformation trends and practices. We are in the process of building a dedicated API in order to provide vetted researchers with access to relevant data on disinformation. We are also in regular engagement with the European Digital Media Observatory. In parallel, over the past months, we have been working on developing a global and separate API that will provide selected researchers with access to various public and anonymised data from our Platform.

Empowering the Fact-Checking Community

TikTok recognises the important contribution of fact-checkers to the fight against disinformation. TikTok's approach differs from that of other platforms, as we incorporate fact-checker input into our broader content moderation efforts. In particular, this means that this feedback is relayed to TikTok's moderation teams so that they can ensure it is factored into their moderation work. This approach effectively produces a force multiplier to the underlying fact-checking output, ensuring that the disinformation content or trend is more comprehensively and broadly addressed.

TikTok has begun to roll out its fact-checking programme in the EU and EEA and we currently work with eight fact-checking partners that review content in ten official languages. All of our fact-checking partners are accredited by the International Fact-Checking Network. While content is being fact-checked or when content can't be substantiated through fact-checking, we may significantly reduce the content's distribution so that fewer people see it. Fact-checkers ultimately do not take action on the content directly. The moderator will instead take into account the fact-checkers' feedback on the accuracy of the content when deciding whether the content violates our Community Guidelines and what action to take. In the last year we have built a database of claims that have previously been fact-checked by our fact-checking partners.

We are continuing to expand our fact-checking program both in terms of the EU and EEA countries in which we have fact-checking partners and scaling up the volume of claims we fact-check and the content our current partners are reviewing. For 2023, we are focusing on setting up further fact-checking partnerships, starting with partners in Portugal, Denmark, Greece and Belgium. As we build momentum with our fact-checking partners, the metrics will also reflect this progress. We also look forward to working with the newly established European Fact-Checking Standards Network.

Looking forward

TikTok is proud to be part of this major new step forward in tackling disinformation. We recognise that collaboration across the ecosystem is critical to eradicating the conditions for harmful disinformation and deceptive behaviour to flourish on platforms. We will continue to collaborate with industry and other partners through the Taskforce and Code subgroups, including (but not limited to): exchanging information on cross platform covert influence operations (commitment 16); working to set standards for vetting and verifying researchers (commitment 27); and agreeing and implementing further structural indicators. We also look forward to making the new Transparency Centre, which we helped co-create, a vital resource for our community, advertisers, regulators, policymakers and governments across Europe.

We look forward to continuing to develop and enhance our policies and tools ahead of the next Code report in July 2023

Definitions	
Definitions *NB: This table sets out the meaning of terms which are defined throughout the report. Certain chapter specific terms are defined within the relevant chapter.*	
CGs	[Community Guidelines]().
CIO	Covert influence operations.
Code	Code of Practice on Disinformation.
DSA	Digital Services Act.
I&A policies	[Integrity and Authenticity policies]().
Misinformation	At TikTok, we take action to moderate or remove content containing "harmful misinformation", irrespective of intent (i.e. whether or not the content in question amounts to disinformation), in accordance with our CGs. However, for the purposes of alignment with the Code, throughout this report, we consider "misinformation" to have effectively the same meaning as "disinformation".
WHO	World Health Organisation.

II. Scrutiny of Ad Placements[1]	
Commitment 1	
Relevant signatories participating in ad placements commit to defund the dissemination of disinformation, and improve the policies and systems which determine the eligibility of content to be monetised, the controls for monetisation and ad placement, and the data to report on the accuracy and effectiveness of controls and services around ad placements	
In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	Yes
If yes, list these implementation measures here [short bullet points].	- Introduced restrictions which prohibit accounts belonging to politicians and political parties from accessing advertising features (although exemptions are made for governments in certain circumstances e.g., for public health). - Partnered with third party Zefr to allow advertisers to obtain campaign insights into brand suitability and safety on the platform. Zefr is fully aligned with GARM's Brand Safety Floor and Suitability Framework (the "**GARM Framework**") and is working on providing additional TikTok measurements related to misinformation.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	- Developing and designing new advertising policies targeted at misinformation. - Partnering with DoubleVerify to provide advertisers with media quality measurement for ads. DoubleVerify is working actively with us to expand their suite of brand suitability and media quality solutions on the Platform.
Measure 1.1	
QRE 1.1.1	To help keep our platform welcoming and authentic for everyone, we take seriously ensuring it is free from harmful misinformation.

[1] Please note that defined terms have the meaning set out in our definitions section here.

(I) Our policies and approach

Our I&A policies within our CGs are the first line of defence in combating harmful misinformation and deceptive behaviours on our platform. All users are required to comply with these CGs. Where content on our platform is deemed violative of our CGs, we remove it.

Paid ads are also subject to our strict underlined advertising policies and are reviewed against these policies before being allowed on our platform. Our ad policies specifically prohibit misleading, inauthentic and deceptive behaviours. This would include unwarranted or deceptive claims, or other misinformation. Ads deemed in violation of these policies will not be permitted on our platform, and accounts deemed in severe or repeated violation may be suspended or banned.

There are also policies in place (e.g. regarding medical claims or certain financial services) under which we check whether advertisers making certain claims or offering certain services which are susceptible to disinformation have the appropriate authority to do so. For example, where an advertiser makes a medical claim, we check whether the advertiser has a marketing authorisation for the claim or for the product or service being advertised.

We have also developed specific advertising policies that focus on certain topics where the risk of disinformation may be higher, with a view to taking a targeted approach to defund the spread of disinformation.

Our Covid-19 advertising policy[2] prohibits ads which present Covid-19 in a distasteful manner e.g., manipulating consumers' fear or anxiety, or spreading harmful misinformation to push sales. As well as ensuring ads relating to Covid-19 do not spread harmful misinformation, we also promoted authoritative sources of information. We provided free ad credits to health authorities, governments, and non-profits to increase vaccine adoption, debunk vaccine myths and promote trustworthy Covid-19 resources.

In line with our approach of building a platform that brings people together, not divides them, we have long prohibited political ads and branded content. Specifically, we do not allow paid ads (nor landing pages) that promote or oppose a candidate, current leader, political party or group, or content that advocates a stance (for or against) on a local, state, or federal issue of public importance in order to influence a political decision or outcome. Similar rules apply in respect of branded content. We also do not allow political actors to place ads[3] and we recently introduced restrictions at an account level. This means accounts belonging to politicians and political parties will automatically have their access to advertising features turned off. As mentioned in the context of Covid-19, we do make exceptions for governments in certain circumstances e.g., to promote public health.

We make various brand safety tools available to advertisers to assist in helping to ensure that their ads are not placed adjacent to content they do not consider to fit with their brand image. While any content that is violative of our CGs, including our I&A policies, is removed, the brand safety tools are designed to help advertisers to further protect their brand; for example, by choosing content less susceptible to disinformation to be near their ad. For example, a family-oriented brand may not want to appear next to videos containing mild political content. As a GARM member we are committed to using and upholding the GARM Framework to do this.

(II) Verification in the context of ads

---

[2] See section 3.8 of the TikTok Advertising Policies
[3] See the 'Political and Issue Based Advertising' section

| | Our strict ad policies require advertisers to meet a number of requirements regarding the landing page. For example, the landing page must be functioning and must contain complete and accurate information including about the advertiser. Ads risk not being approved if the product or service advertised on the landing page does not match that included in the ad.

We also provide verified badges on some accounts including certain advertisers. Verified badges help users make informed choices about the accounts they choose to follow. It's an easy way for notable figures to let users know they're seeing authentic content, and it helps to build trust among high-profile accounts and their followers. For individuals, non-profits, institutions, businesses, or official brand pages, this badge builds an important layer of clarity with the TikTok community. We consider a number of factors before granting a verified badge, such as whether the notable account is authentic, unique, and active.

We are currently trialling mandatory verification for accounts belonging to a government, politician, or political party in the US. Meanwhile, in the EU, verification is available for these types of accounts. We also apply a number of policies that help prevent misuse of certain features e.g., access to advertising features and solicitation for campaign fundraising are not allowed. |
|---|---|

Methodology of data measurement:

We have set out the number of ads that have been removed from our platform for violation of our Covid-19 misinformation, political content and misleading, inauthentic and deceptive behaviours policies respectively. Note that numbers have only been provided for monetised markets and are based on where the ads were displayed.

SLI 1.1.1 – Numbers by actions enforcing policies above (specify if at page and/or domain level)	Number of ad removals under the Covid-19 misinformation ad policy		Number of ad removals under the political content ad policy		Number of ad removals under misleading, inauthentic and deceptive behaviours		N/A	
Level	Page	Domain	Page	Domain	Page	Domain	N/A	N/A
Member States								
Austria	0		2		168			
Belgium	0		1		135			
Bulgaria	0		0		0			
Croatia	0		0		0			
Cyprus	0		0		0			

Czech Republic	0	17	97
Denmark	0	27	95
Estonia	0	0	0
Finland	1	3	60
France	0	19	611
Germany	0	9	1187
Greece	0	0	0
Hungary	0	3	75
Ireland	2	0	508
Italy	1	34	425
Latvia	0	0	0
Lithuania	0	0	0
Luxembourg	0	0	0
Malta	0	0	0
Netherlands	0	2	189
Poland	1	20	211
Portugal	0	2	203
Romania	0	14	636
Slovakia	0	0	0
Slovenia	0	0	0

Spain	1	27	956	
Sweden	0	4	181	
Iceland	0	0	0	
Liechtenstein	0	0	0	
Norway	0	7	65	
Total EU	6	184	5737	
Total EEA	6	191	5802	
This additional Service Level Indicator provides an estimated financial value of the actions taken by Signatories to demonetise disinformation sources (under SLI 1.1.1). It is based on media metrics available to Signatories (query/bid or impression) and applying an agreed-upon conversion factor provided by a third party designated by the Task-force of the Code (Ebiquity plc.).				
SLI 1.1.2 - Preventing the flow of legitimate advertising investment to sites or content that are designated as disinformation	We do not currently offer ad revenue sharing in Europe. We offer creator monetisation opportunities such as the TikTok Creator Fund, which creators meeting certain eligibility criteria are able to join in order to monetise their content based on the level of user interaction with the videos they post. All creators must comply with TikTok's CGs, including our I&A policies. Where creators fail to comply, they will be unable to monetise the content. Given the range of potential policy violations that could be engaged across creator monetisation tools, we are currently unable to provide the financial value of the demonetisation of these sources as a result of disinformation. We are working towards being able to provide this, and are continuing to engage with the sub-group and work with the third party engaged by the European Commission (Ebiquity PLC).			
	N/A			
Member States				
List actions per member states and languages (see example table above)				
Measure 1.2.				
QRE 1.2.1	See explanation provided for SLI 1.1.2.			

	We also prohibit accounts verified as belonging to a government, politician or political party from accessing monetization features. Specifically, they will not have access to features like 'tipping', and 'receiving diamonds', and will be ineligible for our Creator Fund. These changes, along with our existing ban on political advertising, mean that accounts belonging to politicians, political parties and governments will not be able to give or receive money through TikTok's monetization features, or spend money promoting their content (although exemptions are made for governments in certain circumstances e.g., for public health).			
SLI 1.2.1	**Methodology of data measurement:** See explanation provided for SLI 1.1.2. Given the range of potential policy violations that could be engaged across creator monetisation tools, we are currently unable to provide a breakdown of the instances in which accounts are prevented from monetising content as a result of disinformation.			
	N/A	N/A	N/A	N/A
Member States				
List actions per member states and languages (see example table above)				
Measure 1.3				
QRE 1.3.1	We partner with a number of industry leaders to provide a number of controls and transparency tools to advertising buyers with regard to the placement of ads: **Controls**: We offer pre-campaign solutions to advertisers so they can put additional safeguards in place before their campaign goes live. These measures are in addition to the CGs and will apply to any content that appears in the For You feed (since this can comprise both ads and other types of content): • **TikTok Inventory Filter:** This is our proprietary system which enables advertisers to choose the content they want their ads to run adjacent to. The Inventory Filter is embedded directly in TikTok Ads Manager, the system through which advertisers purchase ads, and informed by the GARM Framework. More details can be found here. The Inventory Filter policies include topics which may be susceptible to disinformation. • **TikTok Brand Safety Verified by DoubleVerify (previously OpenSlate):** When placing an ad, advertisers can select to ensure their ads run near content verified by DoubleVerify. This setting is intended to prevent certain categories of content from running adjacent to their ads on the platform, including categories, such as Politics and Sensitive Issues, which may be particularly susceptible to disinformation. • **TikTok Brand Safety by Integral Ad Science ("IAS")**: Advertisers can select IAS within the platform to ensure their ads run near content verified by IAS. IAS verifies content against the GARM Framework. Some disinformation content may be filtered by IAS as			

	a result of the existing GARM Framework categories, such as Sensitive Social Issues. IAS is exploring including a category of "misinformation" in the future.

Transparency: We have partnered with third parties to offer post-campaign solutions that enable advertisers to assess the suitability of user content that ran immediately adjacent to their ad in the For You feed:

- **Zefr**: Through our partnership with Zefr, advertisers can obtain campaign insights into brand suitability and safety on the platform. Zefr is fully aligned with the GARM Framework and is working on providing additional TikTok measurements related to misinformation. Zefr also recently acquired the AI misinformation company Adverif.ai which is powered by fact-checking data from more than 50 organisations globally.
- **IAS**: Advertisers can measure brand safety, viewability and invalid traffic on the platform with the IAS Signal platform. As with IAS's pre-bid solution covered above, IAS is exploring including a category of "misinformation" in the future.
- **DoubleVerify**: We are partnering with DoubleVerify to provide advertisers with media quality measurement for ads. DoubleVerify is working actively with us to expand their suite of brand suitability and media quality solutions on the platform. |
| Measure 1.4 | |
| **QRE 1.4.1** | When TikTok advertises, we buy advertising space only through ad networks (either directly or through agencies) which make available to us robust and reputable brand safety tools, which mitigate the risk of TikTok ads appearing next to sources of disinformation. These include:

- **IAS:** As well as making IAS available on the platform, we make use of IAS to ensure our own ads run on or near suitable content. IAS reports on content which has been filtered or flagged and allows us to make changes to the context filters to avoid ads appearing next to similar content in the future.
- **Newsguard**: We also make use of Newguard which uses topic exclusion lists to prevent our ads from being published on sites which feature harmful misinformation on a repeated basis.

We monitor the placement of our ads closely and in the event our ads are found to run adjacent to or on sources of disinformation, we will investigate to determine whether the content in question has been correctly identified and verify this with IAS, Newsguard or GARM as appropriate. We will then adjust any filters as appropriate or add the publication in question to a blocklist to prevent a recurrence.
When we purchase advertising space through ad networks, we make use of IAS in the way set out above, in addition to various brand safety tools, namely Moat, DV and Peer 39, to prevent TikTok ads appearing next to sources of disinformation. |
| Measure 1.5 | |
| **QRE 1.5.1** | We have achieved the TAG Brand Safety Certified seal by the Trustworthy Accountability Group ("**TAG**") in Europe and globally. We have been certified by the Internet Advertising Bureau ("**IAB**") for the IAB UK Gold Standard 2.1.

We will also be complying with our independent audit obligations under the DSA. |
| **QRE 1.5.2** | As set out in QRE 1.5.1 above we have achieved the TAG Brand Safety Certified seal and the IAB UK Gold Standard 2.1. |

Measure 1.6	
QRE 1.6.1	Zefr (one of the brand safety tools we offer) is working on providing us with additional measurements related to misinformation. Zefr recently acquired the AI misinformation company Adverif.ai, which is powered by fact-checking data from more than 50 organisations globally. We are also exploring what other partnerships, including with fact-checkers and researchers, may be appropriate for our platform. Furthermore, our strict advertising policies help to ensure that the categories of content which are most likely to require such checks and integration of information do not make it onto the platform in the first place.
QRE 1.6.2	As set out in QRE 1.4.1, we only purchase ads through ad networks which make available to us robust and reputable brand safety tools. All of our media investment is thus protected by such tools.
QRE 1.6.3	See QRE 1.3.1 for details of post-bid solutions which enable transparency about processes and outcomes.
QRE 1.6.4	Not applicable as TikTok does not rate sources.
SLI 1.6.1	N/A
	As set out in QRE 1.4.1, we only purchase ads through ad networks which make available to us robust and reputable brand safety tools. All of our media investment is thus protected by such tools.
Member States	
List actions per member states and languages (see example table above)	

II. Scrutiny of Ad Placements
Commitment 2
Relevant Signatories participating in advertising commit to prevent the misuse of advertising systems to disseminate Disinformation in the form of advertising messages.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	N/A

Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	● Drafting new misinformation policies and devising a feasible launch plan. ● Expanding on additional tools, methods or partnerships as we work on enhancing our mis- and dis-information advertising policies.

Measure 2.1	
QRE 2.1.1	See QRE 1.1.1 for details of how disinformation is currently addressed in advertising messages. We are in the process of supplementing the existing approach by producing dedicated advertising focused misinformation policies. We intend to provide further information on this initiative in the next report.

SLI 2.1.1 – Numbers by actions enforcing policies above	**Methodology of data measurement:** We have set out the number of ads that have been removed from our platform for violation of our Covid-19 misinformation, political content and misleading, inauthentic and deceptive behaviours policies respectively. Note that numbers have only been provided for monetised markets and are based on where the ads were displayed.		
	Number of ad removals under the Covid-19 misinformation ad policy	**Number of ad removals under the political content ad policy**	**Number of ad removals under misleading, inauthentic and deceptive behaviours**
Member States			
List actions per member states and languages (see example table above)			
Austria	0	2	168
Belgium	0	1	135
Bulgaria	0	0	0

Croatia	0	0	0
Cyprus	0	0	0
Czech Republic	0	17	97
Denmark	0	27	95
Estonia	0	0	0
Finland	1	3	60
France	0	19	611
Germany	0	9	1187
Greece	0	0	0
Hungary	0	3	75
Ireland	2	0	508
Italy	1	34	425
Latvia	0	0	0
Lithuania	0	0	0
Luxembourg	0	0	0
Malta	0	0	0
Netherlands	0	2	189
Poland	1	20	211
Portugal	0	2	203
Romania	0	14	636

Slovakia	0	0	0
Slovenia	0	0	0
Spain	1	27	956
Sweden	0	4	181
Iceland	0	0	0
Liechtenstein	0	0	0
Norway	0	7	65
Total EU	6	184	5737
Total EEA	6	191	5802
Measure 2.2			
QRE 2.2.1	In order to identify content and sources that breach our Covid-19 advertising policy, all ads go through moderation prior to going live on the platform. After the ad goes "live" on the platform, users can report any concerns using the "report" button, and the ad will be reviewed again and appropriate action taken if necessary. TikTok also operates a "recall" process whereby ads already on TikTok will go through an additional stage of review if certain conditions are met, including reaching certain impression thresholds. TikTok also conducts additional reviews on random samples of ads to ensure its processes are functioning as expected. We will explore additional tools, methods or partnerships as we work on enhancing our disinformation advertising policies.		
Measure 2.3			
QRE 2.3.1	See QRE 2.2.1 above.		
SLI 2.3.1.	See SLI 1.1.1 and 2.1.1 above.		
	N/A	N/A	
Member States			

List actions per member states and languages (see example table above)		
Measure 2.4		
QRE 2.4.1	Where an advertiser has violated an advertising policy they are informed by way of a notification. This is visible in their TikTok Ads Manager account or, where an advertiser has booked their ad through a TikTok representative, then the representative will inform the advertiser of any violations. Advertisers are able to make use of functionality to appeal rejections of their ads in certain circumstances. We are currently unable to provide meaningful data in relation to advertiser appeals.	
SLI 2.4.1	N/A	
	N/A	N/A
Member States		
List actions per member states and languages (see example table above)		

II. Scrutiny of Ad Placements
Commitment 3
Relevant Signatories involved in buying, selling and placing digital advertising commit to exchange best practices and strengthen cooperation with relevant players, expanding to organisations active in the online monetisation value chain, such as online e-payment services, e-commerce platforms and relevant crowd-funding/donation systems, with the aim to increase the effectiveness of scrutiny of ad placements on their own services.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No

If yes, list these implementation measures here [short bullet points].	N/A
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No
If yes, which further implementation measures do you plan to put in place in the next 6 months?	N/A
Measure 3.1	
QRE 3.1.1	As set out later on in this report, we cooperate with a number of third parties to facilitate the flow of information that may be relevant for tackling purveyors of harmful misinformation. This information is shared internally to help ensure consistency of approach across our platform. Specifically, in the area of the "online monetisation value chain', in anticipation of COPD requirements to tackle the dissemination of disinformation in advertising and to learn more about fact-checking, our Monetization Integrity team participated in GlobalFact 2022, hosted by Poynter's International Fact-Checking Network ("**IFCN**"). We have also been actively involved in the Task-force subgroup for Chapter 2, working with other signatories to define and outline metrics regarding the monetary reach and impact of harmful misinformation. We are in close collaboration with industry to ensure alignment and clarity on the reporting of these code requirements.
Measure 3.2	
QRE 3.2.1	In addition to what we have set out in relation to QRE 3.1.1, we are a GARM member, and together with our fellow members and signatories, are committed to removing harmful misinformation from monetisation as part of our commitment to upholding the GARM Framework. The misinformation guidelines have been developed in coordination with the European Commission and in consultation with NGO partners, such as Consumers International and Reporters without Borders. GARM members are also working together to share learnings and best practices to achieve this goal, with misinformation the focus of one of the GARM working groups.
Measure 3.3	
QRE 3.3.1	We will explore additional tools, methods or partnerships as we work on enhancing our misinformation advertising policies.

III. Political Advertising[4]	
Commitment 4	
Relevant Signatories commit to adopt a common definition of "political and issue advertising".	
In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	N/A
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No
If yes, which further implementation measures do you plan to put in place in the next 6 months?	N/A
Measure 4.1	
Measure 4.2	
QRE 4.1.1 (for measures 4.1 and 4.2)	We prohibit political advertising, which is acknowledged to be a source of disinformation, with this Code dedicating a full Chapter to the topic of Political Advertising. Specifically, per our advertising policies, we do not allow political actors to place advertising, nor do we allow ads and landing pages which: "*reference, promote or oppose a candidate for public office, current or former political leader, political party, or political organisation. They must not contain content that advocates a stance (for or against) on a local, state, or federal issue of public importance in order to influence a political outcome.*" We allow cause-based advertising and public services advertising from government agencies, non-profits and other entities if they are not driven by partisan political motives.

---

[4] Please note that defined terms have the meaning set out in our definitions section here.

	We also prohibit <u>political content in branded content</u> i.e. content which has been incentivised by a third party and needs to be disclosed in accordance with local advertising and consumer laws.
	We continue to monitor the evolution of the draft Regulation on Transparency of Political Advertising and will carefully consider any further changes to our policies this might require.
QRE 4.1.2 (for measures 4.1 and 4.2)	Not applicable at this stage.

III. Political Advertising
Commitment 5
Relevant Signatories commit to apply a consistent approach across political and issue advertising on their services and to clearly indicate in their advertising policies the extent to which such advertising is permitted or prohibited on their services.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	N/A
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No
If yes, which further implementation measures do you plan to put in place in the next 6 months?	N/A
Measure 5.1	
QRE 5.1.1	This measure does not apply to TikTok for the reasons set out at QRE 4.1.1 above. Should our policies change or an additional definition of "issue-based advertising" be established by the Task-force which includes ads permitted to be served on the platform, then we will report on its progress in relation to each of the Measures and the associated QREs and SLIs.

III. Political Advertising	
Commitment 6	
Relevant Signatories commit to make political or issue ads clearly labelled and distinguishable as paid-for content in a way that allows users to understand that the content displayed contains political or issue advertising	
In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	N/A
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No
If yes, which further implementation measures do you plan to put in place in the next 6 months?	N/A

Measure 6.1	
QRE 6.1.1	See response to QRE 5.1.1
Measure 6.2	
QRE 6.2.1	Whilst we do not consider this provision to apply to TikTok for the reasons set out at QRE 4.1.1 above, TikTok does already provide enhanced information attached to each ad served. The "About this Ad" functionality is accessible from each ad on the platform and enables users to see factors used to deliver that ad to them. This will be further refined in the coming months to ensure compliance with our DSA transparency obligations.
QRE 6.2.2	See response to QRE 5.1.1.
SLI 6.2.1 – numbers for actions enforcing policies above	**N/A**

	N/A	N/A	N/A	N/A
Member States				
List actions per member states and languages (see example table above)				
Measure 6.3				
QRE 6.3.1	See response to QRE 5.1.1.			
Measure 6.4				
QRE 6.4.1	See response to QRE 5.1.1.			
Measure 6.5				
QRE 6.5.1	**Not committed**. This commitment is not applicable as TikTok is not a messaging app.			

III. Political Advertising
Commitment 7
Relevant Signatories commit to put proportionate and appropriate identity verification systems in place for sponsors and providers of advertising services acting on behalf of sponsors placing political or issue ads. Relevant signatories will make sure that labelling and user-facing transparency requirements are met before allowing placement of such ads.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	N/A

Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No	
If yes, which further implementation measures do you plan to put in place in the next 6 months?	N/A	
Measure 7.1		
QRE 7.1.1	See response to QRE 5.1.1.	
SLI 7.1.1 – numbers for actions enforcing policies above (comparable metrics as for SLI 6.2.1)	N/A	
	N/A	N/A
Member States		
List actions per member states and languages (see example table above)		

Measure 7.2	
QRE 7.2.1	See response to QRE 5.1.1.
QRE 7.2.2	See response to QRE 5.1.1.
Measure 7.3	
QRE 7.3.1	See response to QRE 5.1.1.
QRE 7.3.2	See response to QRE 5.1.1.
Measure 7.4	
QRE 7.4.1	See response to QRE 5.1.1.

III. Political Advertising

Commitment 8
Relevant Signatories commit to provide transparency information to users about the political or issue ads they see on their service.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	N/A
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No
If yes, which further implementation measures do you plan to put in place in the next 6 months?	N/A
Measure 8.1	
Measure 8.2	
QRE 8.2.1 (for measures 8.1 & 8.2)	See responses to QRE 5.1.1 and QRE 6.2.1.

III. Political Advertising

Commitment 9
Relevant Signatories commit to provide users with clear, comprehensible, comprehensive information about why they are seeing a political or issue ad.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	N/A
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No
If yes, which further implementation measures do you plan to put in place in the next 6 months?	N/A
Measure 9.1	
Measure 9.2	
QRE 9.2.1 (for measures 9.1 & 9.2)	See response to QRE 6.2.1.

III. Political Advertising
Commitment 10
Relevant Signatories commit to maintain repositories of political or issue advertising and ensure their currentness, completeness, usability and quality, such that they contain all political and issue advertising served, along with the necessary information to comply with their legal obligations and with transparency commitments under this Code.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	N/A

Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Launching the Commercial Content Library, which will include ads and branded content, to comply with our obligations under the DSA.
Measure 10.1	
Measure 10.2	
QRE 10.2.1 (for measures 10.1 & 10.2)	Whilst we do not consider this provision to apply to TikTok for the reasons set out at QRE 4.1.1 above, we will be launching the Commercial Content Library, which will include ads and branded content, to comply with our obligations under the DSA.

III. Political Advertising
Commitment 11
Relevant Signatories commit to provide application programming interfaces (APIs) or other interfaces enabling users and researchers to perform customised searches within their ad repositories of political or issue advertising and to include a set of minimum functionalities as well as a set of minimum search criteria for the application of APIs or other interfaces."

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	N/A
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No

If yes, which further implementation measures do you plan to put in place in the next 6 months?	N/A
Measure 11.1	
Measure 11.2	
Measure 11.3	
Measure 11.4	
QRE 11.1.1 (for measures 11.1-11.4)	Whilst we do not consider this provision to apply to TikTok for the reasons set out at QRE 4.1.1 above, we will be launching the Commercial Content Library which will include ads and branded content, to comply with our obligations under the DSA. See response to QRE 5.1.1.
QRE 11.4.1	See response to QRE 5.1.1.

III. Political Advertising
Commitment 12 Relevant Signatories commit to increase oversight of political and issue advertising and constructively assist, as appropriate, in the creation, implementation and improvement of political or issue advertising policies and practices.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	TikTok did not subscribe to this commitment. Commitment 12 and associated measures are applicable to Civil Society only.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No

If yes, which further implementation measures do you plan to put in place in the next 6 months?	TikTok did not subscribe to this commitment. Commitment 12 and associated measures are applicable to Civil Society only.
Measure 12.1	
Measure 12.2	
Measure 12.3	
QRE 12.1.1 (for measures 12.1-12.3)	Commitment 12 and measures 12.1, 12.2 and 12.3 are applicable to Civil Society only.

III. Political Advertising

Commitment 13 Relevant Signatories agree to engage in ongoing monitoring and research to understand and respond to risks related to Disinformation in political or issue advertising.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	N/A
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No
If yes, which further implementation measures do you plan to put in place in the next 6 months?	N/A
Measure 13.1	

Measure 13.2	
Measure 13.3	
QRE 13.1.1 (for measures 13.1-13.3)	See response to QRE 5.1.1. We are committed to remaining engaged with all discussions being held through the Task-force and other fora to ensure our policies and processes remain current and emerging and novel threats are addressed in our policies and enforcement.

IV. Integrity of Services[5]

Commitment 14

In order to limit impermissible manipulative behaviours and practices across their services, Relevant Signatories commit to put in place or further bolster policies to address both misinformation and disinformation across their services, and to agree on a cross-service understanding of manipulative behaviours, actors and practices not permitted on their services. Such behaviours and practices, which should periodically be reviewed in light with the latest evidence on the conducts and TTPs employed by malicious actors, such as the AMITT Disinformation Tactics, Techniques and Procedures Framework, include:

The following TTPs pertain to the creation of assets for the purpose of a disinformation campaign, and to ways to make these assets seem credible:

1. Creation of inauthentic accounts or botnets (which may include automated, partially automated, or non-automated accounts)
2. Use of fake / inauthentic reactions (e.g. likes, up votes, comments)
3. Use of fake followers or subscribers
4. Creation of inauthentic pages, groups, chat groups, fora, or domains
5. Account hijacking or impersonation

The following TTPs pertain to the dissemination of content created in the context of a disinformation campaign, which may or may not include some forms of targeting or attempting to silence opposing views. Relevant TTPs include:

6. Deliberately targeting vulnerable recipients (e.g. via personalised advertising, location spoofing or obfuscation)
7. Deploy deceptive manipulated media (e.g. "deep fakes", "cheap fakes"...)
8. Use "hack and leak" operation (which may or may not include doctored content)
9. Inauthentic coordination of content creation or amplification, including attempts to deceive/manipulate platforms algorithms (e.g. keyword stuffing or inauthentic posting/reposting designed to mislead people about popularity of content, including by influencers)
10. Use of deceptive practices to deceive/manipulate platform algorithms, such as to create, amplify or hijack hashtags, data voids, filter bubbles, or echo chambers
11. Non-transparent compensated messages or promotions by influencers
12. Coordinated mass reporting of non-violative opposing content or accounts

---

[5] Please note that defined terms have the meaning set out in our definitions section here.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	Yes
If yes, list these implementation measures here [short bullet points].	- Increased transparency into our work on CIO and published (here) insights into the networks we identify and remove from our platform globally. - Expanded the roll out of our branded content toggle to all EEA users.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Conducting an in-depth analysis of the TTPs agreed upon at Task-force level and where needed, expanding existing or developing additional policies as a result.
Measure 14.1	
QRE 14.1.1	As well as our I&A policies in our CGs which safeguard against harmful misinformation (see QRE 18.2.1), our integrity policies also robustly prohibit deceptive behaviours. Our policies on deceptive behaviours relate to the TTPs as follows: *TTPs which pertain to the creation of assets for the purpose of a disinformation campaign, and to ways to make these assets seem credible:* **Creation of inauthentic accounts or botnets (which may include automated, partially automated, or non-automated accounts)** Our platform security policies under our CGs expressly prohibit any attempt to undermine or abuse the security, integrity or reliability of our platform, products, or services. This includes leveraging TikTok accounts under false or fraudulent pretences to distribute spam, phishing, or smishing content in an attempt to perpetrate cybercrime or gain unauthorised access to others' content, accounts, systems, or data. **Use of fake / inauthentic reactions (e.g. likes, up votes, comments) and use of fake followers** Our integrity policies which address **fake engagement** prohibit any content or activity that seeks to artificially inflate popularity on our platform and any attempts to manipulate the platform to increase interaction metrics. We do not allow our users to: - share instructions on how to artificially increase views, likes, followers, shares, or comments; - engage in selling or buying views, likes, followers, shares, or comments;

- promote artificial traffic generation services; or
- create malicious software or modify code to artificially increase views, likes, followers, shares, or comments.

As a result of identifying such repeated activity on our platform, we may deem the videos or the account to be spam and remove them.

Creation of inauthentic pages, groups, chat groups, fora, or domains

TikTok does not have pages, groups, chat groups, fora or domains. As such, this TTP is not relevant to the platform.

Account hacking or Impersonation

Our policies prohibit **impersonation** which refers to accounts that pose as another person or entity in a deceptive manner. Our users are not allowed to use someone else's name, biographical details, or profile picture in a misleading manner. In order to protect freedom of expression, we do allow accounts that are clearly parody, commentary, or fan-based, such as where the username indicates that it is a fan, commentary, or parody account and not affiliated with the subject of the account.

Our platform security policies under our CGs also expressly prohibit users from providing access to their account credentials to others or enable others to conduct activities against our CGs. We prohibit use of automated scripts, web crawling, software, deceptive techniques, or any other way to attempt to obtain, acquire, or request login credentials or other sensitive information from TikTok or its users. In addition, to ensure account security, we do not allow users to click on suspicious links or engage in requests for information about your TikTok account details, passwords, verification qualification, financial, or other personal information.

TTPs which pertain to the dissemination of content created in the context of a disinformation campaign, which may or may not include some forms of targeting or attempting to silence opposing views:

Deliberately targeting vulnerable recipients (e.g. via personalised advertising, location spoofing or obfuscation)

We fight against **CIO as our policies** prohibit organised attempts to manipulate or corrupt public debate while also misleading our systems or users about the identity, origin, operating location, popularity or overall purpose.

Deceptive manipulated media (e.g. "deep fakes", "cheap fakes"...)

Our policies prohibit **deceptive synthetic and manipulated media**:

	- Synthetic media refers to content that uses editing tools or technology to deceive viewers by showing a version of events that has been substantially altered or never occurred. We remove content created or modified using artificial intelligence ("**AI**") or deep learning technology that (i) misleads about the truth of events, and (ii) could be reasonably expected to cause harm. - Manipulated media refers to content that deceives or misleads people by altering content without necessarily using any AI (e.g. altering the frames of a video — cropping, changing speed, or adding or deleting visual information — and/or editing together disparate media) and, by doing so, alters how a viewer would perceive a story. **Non-transparent compensated messages or promotions by influencers** Our Terms of Service and Branded Content Policy require users posting about a **brand or product in return for any payment or other incentive** to disclose their content by enabling the branded content toggle which we make available for users. We also provide functionality to enable users to report suspected undisclosed branded content which reminds the user who posted the suspected undisclosed branded content of our requirements and prompts them to turn the branded content toggle on if required.
QRE 14.1.2	At TikTok, we place considerable emphasis on proactive content moderation and use a combination of technology and safety professionals to detect and remove harmful misinformation (see QRE 18.1.1) and deceptive behaviours on our Platform, *before* it is reported to us by users or third parties. For instance, we take proactive measures to prevent inauthentic or spam accounts from being created. Thus, we have created and use detection models and rule engines that: - prevent inauthentic accounts from being created based on malicious patterns that have been previously identified; and - remove registered accounts based on certain signals (ie, uncommon behaviour on the platform). We also manually monitor user reports of inauthentic accounts in order to detect larger clusters or similar inauthentic behaviours. However, given the complex nature of the TTPs, human moderation is critical to success in this area and TikTok's moderation teams therefore play a key role assessing and addressing identified violations. In addition, where content reaches certain levels of popularity in terms of the number of video views, it will be flagged for further review. Such review is undertaken given the extent of the content's dissemination and the increase in potential harm if the content is found to be in breach of our CGs including our I&A policies. We have also set up specifically-trained teams that are focused on investigating and detecting CIO on our Platform. When we investigate and remove these operations, we focus on behaviour and assessing linkages between accounts and techniques to determine if actors are engaging in a coordinated effort to

	mislead TikTok's systems or our community. In each case, we believe that the people behind these activities coordinate with one another to misrepresent who they are and what they are doing.
Measure 14.2	
QRE 14.2.1	The implementation of our policies is ensured by different means, including specifically-designed tools (such as toggles to disclose branded content - see QRE 14.1.1) or human investigations to detect deceptive behaviours (for covert influence activities - see QRE 14.1.2). The implementation of these policies is also ensured through enforcement measures applied in all Member States. Where our teams have a high degree of confidence that an account amounts to an **impersonation**, or is engaged in or is connected to networks we took down in the past as part of a **CIO**, it is removed from our Platform. Please note that in relation to CIO, we are only able to provide the Q3 2022. We are building and testing data infrastructure that can provide information requested at a high level of fidelity. Additionally, CIO investigations are highly resource heavy and require in-depth analysis to ensure high confidence in proposed actions. Similarly, where our teams have a high degree of confidence that a specific content violates one of our TTPs-related policies (See QRE 14.1.1), such content is removed from TikTok. Lastly, we may reduce the discoverability of some content, including by making videos ineligible for recommendation in the For You feed section of our platform. This is, for example, the case for content that tricks or manipulates users in order to inauthentically increase followers, likes, or views.

SLI 14.2.1 – SLI 14.2.4	

TTP OR ACTION	**TTP No. 1:** Creation of inauthentic accounts or botnets (which may include automated, partially automated, or non-automated accounts) **Methodology of data measurement** We have based the number of: (i) fake accounts removed; and (ii) followers of the fake accounts (identified at the time of removal of the fake account), on the country the fake account was last active in.

	SLI 14.2.1		SLI 14.2.2			SLI 14.2.3			SLI 14.2.4			
	N/A	Nr of actions taken by type	N/A	Interaction/ engagement before action	N/A	N/A	N/A	N/A	N/A	TTPs related content in relation to overall content on the service	N/A	N/A
Member States												
List actions per member states (see example table above)		**Number of fake accounts removed**		**Number of followers of fake accounts identified at the time of removal**						**Fake accounts as a % of monthly active users**		
Austria		6774		355360								
Belgium		13321		2935797								
Bulgaria		7602		267899								

Croatia		3205		330193								
Cyprus		1144		33914								
Czech Repub lic		11017		441055								
Denmark		10389		263741								
Estonia		3917		22827								
Finland		11653		55705								
France		175636		2226422								
Germany		138316		2074960								
Greece		11129		449172								
Hungary		9196		176826								

Ireland		13255		647092								
Italy		121962		1110990								
Latvia		4951		96947								
Lithuania		21621		82510								
Luxembourg		3061		98409								
Malta		591		12813								
Netherlands		36653		564114								
Poland		43388		574186								
Portugal		12787		394850								
Romania		20593		2100453								
Slovakia		6453		65156								
Slovenia		3504		23731								

Spain		155220		2268245								
Sweden		17588		453119								
Iceland		2931		14606								
Liechtenstein		319		320								
Norway		7021		210593								
Total EU		864926		18126486					0.0067			
Total EEA		875197		18352005								

TTP OR ACTION	**TTP no. 2:** Use of fake / inauthentic reactions (e.g. likes, up votes, comments)			
	Methodology of data measurement:			
	We based the number of fake likes that we removed on the country of registration of the user. We also based the number of fake likes prevented on the country of registration of the user.			
	SLI 14.2.1	**SLI 14.2.2**	**SLI 14.2.3**	**SLI 14.2.4**

	N/A	Nr of actions taken by type	N/A	Interaction/ engagement before action	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Member States												
List actions per member states (see example table above)		**Number of fake likes removed**		**Number of fake likes prevented**								
Austria		803718		7059247								
Belgium		654348		9674029								
Bulgaria		75224		4600356								
Croatia		52874		2072142								

Cyprus		127097		1900522								
Czech Republic		211989		1787290								
Denmark		206923		3986523								
Estonia		34362		530624								
Finland		152027		2964126								
France		2980294		47372500								
Germany		6245955		103217673								
Greece		389940		6161517								
Hungary		89173		2575760								
Ireland		139225		2061031								
Italy		2142090		31932140								
Latvia		47411		2412481								

Lithuania		42700		618873							
Luxembourg		105232		3718969							
Malta		160039		714299							
Netherlands		1493330		18566080							
Poland		913653		7915210							
Portugal		207787		2863395							
Romania		361464		5981575							
Slovakia		131616		754932							
Slovenia		39103		1427860							
Spain		1136865		35935997							
Sweden		1187727		8806086							
Iceland		10999		364147							

Liechtenstein		0		2200								
Norway		415658		5795213								
Total EU		20280991		317611239								
Total EEA		20707648		323772799								
TTP OR ACTION	**TTP No. 3:** Use of fake followers or subscribers **Methodology of data measurement:** We based the number of fake followers that we removed on the country of registration of the user. We also based the number of fake followers prevented on the country of registration of the user.											
	SLI 14.2.1		**SLI 14.2.2**				**SLI 14.2.3**			**SLI 14.2.4**		
	N/A	Nr of actions taken by type	N/A	Interaction/ engagement before action	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Member States												

List actions per member states (see example table above)		Number of fake followers removed		Number of fake follows prevented									
Austria		4182016		445387									
Belgium		4663537		294305									
Bulgaria		829090		120064									
Croatia		425275		30712									
Cyprus		423285		31199									
Czech Republic		831665		61868									
Denmark		1619583		280910									
Estonia		258266		104886									
Finland		1271446		143467									
France		23136696		2261797									

Germany		41556845		4192434								
Greece		1414391		238038								
Hungary		510228		73421								
Ireland		931538		229312								
Italy		13644642		1953288								
Latvia		356558		55679								
Lithuania		237728		156865								
Luxembourg		297830		68411								
Malta		141770		16023								
Netherlands		7706366		883911								
Poland		2317680		179517								
Portugal		807943		96968								

Romania		2489586		310485									
Slovakia		193913		102822									
Slovenia		196024		16267									
Spain		6793157		506630									
Sweden		6874422		567712									
Iceland		54534		5210									
Liechtenstein		3574		373									
Norway		2755754		301603									
Total EU		124111480		13422376									
Total EEA		126925342		13729562									
TTP OR ACTION	**TTP No. 5:** Account hijacking or impersonation **Methodology of data measurement:** The number of accounts removed under our impersonation policy is based on the active location of the users.												

	SLI 14.2.1		SLI 14.2.2				SLI 14.2.3			SLI 14.2.4		
	N/A	Nr of actions taken by type	N/A	N/A	N/A	N/A	N/A	N/A	N/A	TTPs related content in relation to overall content on the service	N/A	N/A
Member States		**Number of account banned under impersona tion policy**								**Impersonat ion accounts as a % of monthly active users**		
List actions per member states (see example table above)												
Austria		105										
Belgium		112										

Bulgaria		118										
Croatia		43										
Cyprus		12										
Czech Repub lic		56										
Denmark		64										
Estonia		10										
Finland		78										
France		499										
Germany		841										
Greece		69										
Hungary		57										
Ireland		100										

Italy		358										
Latvia		34										
Lithuania		35										
Luxembourg		1										
Malta		9										
Netherlands		204										
Poland		498										
Portugal		132										
Romania		312										
Slovakia		46										
Slovenia		20										
Spain		314										

Sweden		183										
Iceland		9										
Liechtenstein		0										
Norway		112										
Total EU		4310							0.00003			
Total EEA		4431										

TTP OR ACTION	**TTP No. 6.** Deliberately targeting vulnerable recipients (e.g. via personalised advertising, location spoofing or obfuscation)
	Methodology of data measurement:
	The number of CIO networks targeting EEA countries relates to the Q3 2022 period.
	We assess that the network operated from Russia and targeted European countries, primarily Germany, Italy and the United Kingdom. The individuals behind these accounts created localised inauthentic accounts and shared content in German, Italian and English using speech synthesis in these languages amplifying a pro-Russia viewpoint targeting discourse about the war in Ukraine.
	The number of accounts linked to CIO networks identified and removed are based on the geographic location of network operation and technical and behavioural evidence from proprietary and open sources. The number of followers of CIO networks has been based on the number of accounts that followed any account within a network as of the date of that network's removal.

SLI 14.2.1	**SLI 14.2.2**	**SLI 14.2.3**	**SLI 14.2.4**

	Nr of instances of identified TTPs	Nr of actions taken by type	N/A	Interaction/ engagement before action	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Member States												
List actions per member states (see example table above)	**Number of identified networks**	**Number of accounts in the network**		**Number of followers of the network**								
Total EEA	1	1686		133564								
TTP OR ACTION	**TTP No. 11.** Non-transparent compensated messages or promotions by influencers **Methodology of data measurement:** We have based the number of times the branded content toggle has been used on the active location of the users.											
	SLI 14.2.1	**SLI 14.2.2**			**SLI 14.2.3**			**SLI 14.2.4**				

	N/A	Nr of actions taken by type	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Member States		**Number of times the branded content toggle has been used to disclose the existence of a commercial relationship**										
List actions per member states (see example table above)												
Austria		15256										
Belgium		31153										

Bulgaria		7239										
Croatia		2869										
Cyprus		8420										
Czech Republic		6821										
Denmark		5328										
Estonia		979										
Finland		3900										
France		123082										
Germany		132740										
Greece		41108										
Hungary		7367										
Ireland		8534										

Italy		113283										
Latvia		8101										
Lithuania		2367										
Luxembourg		2520										
Malta		1785										
Netherlands		27114										
Poland		63376										
Portugal		78574										
Romania		32921										
Slovakia		2561										
Slovenia		1226										
Spain		163913										

Sweden		15759											
Iceland		434											
Liechtenstein		10											
Norway		7481											
Total EU		908296											
Total EEA		916221											
Measure 14.3													
QRE 14.3.1			We have been engaging with the Integrity of Services Subgroup to set up the first list of TTPs.										

IV. Integrity of Services
Commitment 15
Relevant Signatories that develop or operate AI systems and that disseminate AI-generated and manipulated content through their services (e.g. deep fakes) commit to take into consideration the transparency obligations and the list of manipulative practices prohibited under the proposal for Artificial Intelligence Act.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No

If yes, list these implementation measures here [short bullet points].	N/A
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Working on updating our current synthetic media policy, and developing a longer-term, comprehensive policy and user education strategy.
Measure 15.1	
QRE 15.1.1	Our policies prohibit deceptive synthetic content. On this basis, we remove any identified content that is created or modified using AI technologies in order to mislead our users about the truth of events and could be reasonably expected to cause harm (See QRE 14.1.1). In addition, we are currently working on evolving our current policies and developing new ones to further address AI manipulative practices.
Measure 15.2	
QRE 15.2.1	We are committed to ensuring the algorithms we develop comply with Measure 15.2 of the Code. To that end: ● We have set up internal guidelines and training to help ensure that the training and deployment of our AI algorithms comply with applicable data protection laws, as well as principles of fairness. ● We have instituted a compliance review process for new AI algorithms that meet certain thresholds, and are working to prioritise review of previously developed algorithms. Our efforts in this area are ongoing, and we expect to be able to report on additional developments to help ensure our AI algorithms comply with Measure 15.2 of the Code, in the next reporting period.

IV. Integrity of Services

Commitment 16
Relevant Signatories commit to operate channels of exchange between their relevant teams in order to proactively share information about cross-platform influence operations, foreign interference in information space and relevant incidents that emerge on their respective services, with the aim of preventing dissemination and resurgence on other services, in full compliance with privacy legislation and with due consideration for security and human rights risks.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	N/A
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No
If yes, which further implementation measures do you plan to put in place in the next 6 months?	N/A

Measure 16.1			
QRE 16.1.1	The cross-platform sharing forum referred to in Commitment 16 has not been set up yet. Whilst we wait for the forum to be set up, we will continue to share relevant insights and metrics within our quarterly transparency reports, which aim to inform industry peers and the research community.		
SLI 16.1.1 – Numbers of actions as a result of information sharing	N/A		
	N/A	N/A	N/A
Member States			
List actions per member states and languages (see example table above)			
Measure 16.2			

QRE 16.2.1	The cross-platform sharing forum referred to in Commitment 16 has not been set up yet.

<table>
<tr><td colspan="2" align="center">V. Empowering Users[6]</td></tr>
<tr><td colspan="2" align="center">Commitment 17</td></tr>
<tr><td colspan="2" align="center">In light of the European Commission's initiatives in the area of media literacy, including the new Digital Education Action Plan, Relevant Signatories commit to continue and strengthen their efforts in the area of media literacy and critical thinking, also with the aim to include vulnerable groups.</td></tr>
<tr><td>In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]</td><td>Yes</td></tr>
<tr><td>If yes, list these implementation measures here [short bullet points].</td><td>Piloted the state-controlled media label policy by applying it on media entities in RU, UA and BY. The policy was then expanded to media entities across 40 countries in <u>January 2023</u> . Users across all EEA countries can view the label when they come across the content or profile pages of labelled entities.
Continued our extensive in-app interventions (including video tags, search interventions and in-app information hubs) around Covid-19, Covid-19 Vaccine, Holocaust Denial, Monkey Pox and the War in Ukraine and rolled them out in 21 official EU languages (plus, for EEA users, Norwegian and, as the spoken language of Liechtenstein, German).</td></tr>
<tr><td>Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]</td><td>Yes</td></tr>
<tr><td>If yes, which further implementation measures do you plan to put in place in the next 6 months?</td><td>Expanding the application of state-controlled media labels to additional countries.
Aiming to work with our fact-checking partners to identify specific disinformation trends in countries and develop tailored, localised media literacy campaigns to tackle those trends.
Notifying creators when their content has been made ineligible for recommendation and enabling them to appeal - this work is being undertaken in the context of TikTok's obligations under the DSA (Articles 17 and 20), with an implementation date during the Summer of 2023.</td></tr>
</table>

---

[6] Please note that defined terms have the meaning set out in our definitions section <u>here</u>.

	• Rolling out three media literacy campaigns in Europe in partnership with trusted organisations, starting with a campaign to address disinformation related to the war on Ukraine in certain Eastern European countries - ultimately aiming to improve the digital literacy of users. Where possible and where appropriate, we will aim to scale these campaigns across Europe later in the year.
Measure 17.1	
QRE 17.1.1	Together with systematically removing content from our platform that violates our integrity policies, we have dedicated significant resources to increasing the number of in-app measures that show users additional context on certain content or redirect them to authoritative information and to making those tools available in 21 EU official languages (plus, for EEA users, Norwegian and, as the spoken language of Liechtenstein, German).
	We work with external experts (e.g. WHO on medical information) and our fact checker partners, and we take into account their feedback, as well as user feedback, in order to continually identify new topics and consider which tools may be best suited for raising awareness around that topic and combating harmful misinformation. For example, recently we have moved from creating our own resource pages on certain topics to partnering with external experts and redirecting users to their resources, meaning users have access to the most up-to-date information at all times.
	Since 2020, on topics such as Covid-19, Covid-19 Vaccine, Holocaust Denial, MonkeyPox and War in Ukraine, we deployed a combination of a number of in-app intervention tools:
	• **video notice tags.** These tags are applied to videos containing words or hashtags for the topic. The tag, which will invite the user to "*Learn more about* [the topic]", is clickable and will redirect the user to a trusted resource page.
	• **search intervention.** If a user searches for keywords related to the topic, the user will be displayed a banner which may encourage them to verify the facts and will provide a link to a resource page. If the search term is violative, the user will not be displayed results and instead redirected to a trusted resource page.
	• **public service announcement.** If a user searches for a hashtag on the topic, the user will be displayed a public service announcement that will remind users about our CG's and present them with links to trusted resource pages.
	• **online and in-app information hubs and safety centre pages**. The above tools will often link to resource page(s) which direct users to accurate and up-to-date information from trusted sources.

Depending on the topic or EU country involved, the user could be directed to an external authoritative source (e.g., a national government website), an in-app information hub (e.g., War on Ukraine) or to a dedicated page on our safety centre website (e.g., Covid-19 and Elections integrity).

As well as the tools we deploy to increase users' digital citizenship regarding specific topics, we also apply labels, irrespective of the topic, to encourage users to consider the reliability of the content or the source. These labels are available to users in 21 EU official languages (plus, for EEA users, Norwegian and, as the spoken language of Liechtenstein, German).

- **Unverified content label.** When sometimes our fact checking partners determine that checks are inconclusive or content is not able to be confirmed by our moderators, which is especially common during unfolding events or crises, we apply a banner to inform users the content has been reviewed but cannot be conclusively validated. The goal is to raise users' awareness about the credibility of the content and to reduce sharing (see screenshots here).
    - The video will become ineligible for recommendation into anyone's For You feed to limit the spread of potentially misleading information.
    - The video's creator is also notified that their video was flagged as unsubstantiated content and is provided additional information about why the warning label has been added to their content. This is again to raise the creator's awareness about the credibility of the content that they have shared.

- **State controlled media label**. In the EU, Iceland and Liechtenstein, we have taken steps to restrict access to content from Russia Today, Sputnik, Rossiya RTR / RTR Planeta, Rossiya 24 / Russia 24 and TV Centre International. We apply a prominent label to all other content or accounts from state affiliated media. The user is also automatically shown a full screen pop-up providing the user with information about what the label means and inviting the user to click on "learn more" and be redirected to an in-app page. The measure brings transparency to our community and raises awareness among users to encourage users to consider the reliability of the source.

    - We started with piloting its use on some accounts (and their content) of media entities in Russia, Ukraine and Belarus. We continue to work with experts to inform our global approach and expansion and have recently expanded its use to media entities across 40 countries.

For the unverified content and state controlled media labels, we do not have engagement data available for Q4 2022 and are working on resolving this ahead of the next reporting period.

In 2021, we launched a redesigned Safety Center (all users received an in-box notification about this). The redesign was informed by consultation with users; we made it more colourful and used more images and

	graphics and created a series of longer length videos to bolster our efforts to communicate with users via a variety of media formats. Relating to combating harmful misinformation, we have pages on:		
	Safety Partners: This section provides details of some of TikTok's work with industry experts, non-governmental organisations, and industry associations around the world in our commitment to building a safe platform for our community. **Specific topics:** With helpful resources such as Covid-19 and Elections integrity. We also use our **Newsroom posts** to communicate with our community transparently and to build and maintain trust. We publish a range of posts in our Newsroom in which we seek, among other things, to generate awareness on safety and content related issues for example, highlight new features like our unverified content label.		
SLI 17.1.1 - actions enforcing policies above	**Methodology of data measurement:** The number of impressions, clicks and click through rates of video notice tags, search interventions and public service announcements are based on the active location of the users that engaged with the tools. The number of impressions of the Safety Center pages is based on the IP location of the users.		
	Total count of the tool's impressions	Interactions/ engagement with the tool	Other relevant metrics
Member States			
List actions per member states and languages (see example table above)	**Number of impressions of Video Notice Tag covered by Intervention (Covid-19)**	**Number of clicks of Video Notice Tag covered by Intervention (Covid-19)**	**Click Through Rate of Video Notice Tag covered by Intervention (Covid-19)**
Austria	22,417,955	19,944	0.09%
Belgium	24,438,233	26,987	0.11%
Bulgaria	10,055,563	8,621	0.09%
Croatia	8,209,571	5,718	0.07%
Cyprus	2,814,092	3,072	0.11%

Czech Republic	12,665,865	15,253	0.12%
Denmark	10,980,041	4,774	0.04%
Estonia	2,612,374	2,740	0.10%
Finland	14,181,540	12,914	0.09%
France	143,473,218	160,429	0.11%
Germany	244,857,489	200,328	0.08%
Greece	25,684,748	25,916	0.10%
Hungary	16,938,075	15,053	0.09%
Ireland	35,885,722	24,897	0.07%
Italy	230,422,264	316,117	0.14%
Latvia	4,768,896	5,305	0.11%
Lithuania	5,169,354	4,861	0.09%
Luxembourg	1,763,322	1,712	0.10%
Malta	1,482,589	1,667	0.11%
Netherlands	44,006,929	46,411	0.11%
Poland	69,197,484	61,942	0.09%
Portugal	20,533,602	24,743	0.12%
Romania	38,852,756	44,880	0.12%
Slovakia	6,131,593	4,686	0.08%
Slovenia	3,981,842	3,753	0.09%

Spain	160,385,873	214,298	0.13%
Sweden	26,310,387	28,421	0.11%
Iceland	1,092,451	1,172	0.11%
Liechtenstein	22,759	53	0.23%
Norway	16,129,117	6,172	0.04%
Total EU	**1,188,221,377**	**1,285,442**	**0.11%**
Total EEA	**1,205,465,704**	**1,292,839**	**0.11%**
Member States			
List actions per member states and languages (see example table above)	**Number of impressions of Video Notice Tag covered by Intervention (Covid-19 Vaccine)**	**Number of clicks of Video Notice Tag covered by Intervention (Covid-19 Vaccine)**	**Click Through Rate of Video Notice Tag covered by Intervention (Covid-19 Vaccine)**
Austria	2,447,110	68	0.00%
Belgium	5,449,142	6,594	0.12%
Bulgaria	1,394,591	94	0.01%
Croatia	286,029	21	0.01%
Cyprus	89,995	13	0.01%
Czech Republic	1,605,901	80	0.00%
Denmark	806,716	26	0.00%
Estonia	47,149	5	0.01%

Finland	581,160	19	0.00%
France	38,247,876	43,680	0.11%
Germany	43,001,287	39,130	0.09%
Greece	2,578,642	51	0.00%
Hungary	1,232,786	36	0.00%
Ireland	8,584,309	4,834	0.06%
Italy	49,606,158	97,275	0.20%
Latvia	179,055	11	0.01%
Lithuania	180,422	17	0.01%
Luxembourg	85,494	18	0.02%
Malta	89,734	18	0.02%
Netherlands	10,774,841	12,874	0.12%
Poland	10,260,699	10,271	0.10%
Portugal	2,620,334	79	0.00%
Romania	5,923,355	138	0.00%
Slovakia	244,303	17	0.01%
Slovenia	82,042	12	0.01%
Spain	34,085,834	55,575	0.16%
Sweden	2,969,633	40	0.00%
Iceland	32,437	8	0.02%

	Number of impressions of Video Notice Tag covered by Intervention (Holocaust Denial)	Number of clicks of Video Notice Tag covered by Intervention (Holocaust Denial)	Click Through Rate of Video Notice Tag covered by Intervention (Holocaust Denial)
Liechtenstein	299	2	0.67%
Norway	907,758	27	0.00%
Total EU	223,454,597	270,996	0.12%
Total EEA	224,395,091	271,033	0.12%
Member States			
List actions per member states and languages (see example table above)	**Number of impressions of Video Notice Tag covered by Intervention (Holocaust Denial)**	**Number of clicks of Video Notice Tag covered by Intervention (Holocaust Denial)**	**Click Through Rate of Video Notice Tag covered by Intervention (Holocaust Denial)**
Austria	2,448,910	3,797	0.16%
Belgium	5,943,640	10,497	0.18%
Bulgaria	625,320	1,498	0.24%
Croatia	14,510	34	0.23%
Cyprus	219,170	802	0.37%
Czech Republic	1,628,410	4,440	0.27%
Denmark	1,666,570	3,108	0.19%
Estonia	5,370	13	0.24%
Finland	2,134,370	4,900	0.23%
France	60,621,820	98,187	0.16%
Germany	18,059,540	27,063	0.15%

Greece	2,050,290	4,921	0.24%
Hungary	2,572,220	7,356	0.29%
Ireland	2,872,980	4,241	0.15%
Italy	8,274,970	20,190	0.24%
Latvia	4,380	9	0.21%
Lithuania	8,140	14	0.17%
Luxembourg	11,620	15	0.13%
Malta	11,730	19	0.16%
Netherlands	6,309,110	10,486	0.17%
Poland	15,872,070	50,768	0.32%
Portugal	1,447,560	3,043	0.21%
Romania	2,692,490	7,342	0.27%
Slovakia	14,270	30	0.21%
Slovenia	6,030	11	0.18%
Spain	4,339,200	13,000	0.30%
Sweden	4,633,990	10,949	0.24%
Iceland	183,970	485	0.26%
Liechtenstein	2,800	2	0.07%
Norway	2,442,860	4,709	0.19%
Total EU	144,488,680	286,733	0.20%

Total EEA	147,118,310	291,929	0.20%
Member States			
List actions per member states and languages (see example table above)	**Number of impressions of Video Notice Tag covered by Intervention (Monkey Pox)**	**Number of clicks of Video Notice Tag covered by Intervention (Monkey Pox)**	**Click Through Rate of Video Notice Tag covered by Intervention (Monkey Pox)**
Austria	350,530	542	0.15%
Belgium	271,110	451	0.17%
Bulgaria	159,190	351	0.22%
Croatia	108,970	159	0.15%
Cyprus	39,110	89	0.23%
Czech Republic	156,070	346	0.22%
Denmark	175,460	299	0.17%
Estonia	47,890	75	0.16%
Finland	214,610	406	0.19%
France	782,030	1,523	0.19%
Germany	3,123,430	4,872	0.16%
Greece	301,960	688	0.23%
Hungary	157,780	271	0.17%
Ireland	246,060	344	0.14%
Italy	580,870	999	0.17%

Latvia	71,180	118	0.17%
Lithuania	105,170	143	0.14%
Luxembourg	22,500	29	0.13%
Malta	19,260	35	0.18%
Netherlands	576,130	851	0.15%
Poland	416,210	796	0.19%
Portugal	168,670	340	0.20%
Romania	841,330	1,421	0.17%
Slovakia	87,780	136	0.15%
Slovenia	406,710	707	0.17%
Spain	1,874,220	2,808	0.15%
Sweden	437,540	757	0.17%
Iceland	16,760	31	0.18%
Liechtenstein	710	0	0.00%
Norway	236,620	394	0.17%
Total EU	11,741,770	19,556	0.17%
Total EEA	11,995,860	19,981	0.17%
List actions per member states and languages (see example table above)	**Number of impressions of Search interventions (Covid-19)**	**Number of clicks of Search interventions (Covid-19)**	**Click Through Rate of Search interventions (Covid-19)**

Austria	7,287	116	1.59%
Belgium	15,783	218	1.38%
Bulgaria	159	0	0.00%
Croatia	108	1	0.93%
Cyprus	129	1	0.78%
Czech Republic	152	3	1.97%
Denmark	191	2	1.05%
Estonia	25	0	0.00%
Finland	10,202	127	1.24%
France	155,744	2,258	1.45%
Germany	67,930	923	1.36%
Greece	205	0	0.00%
Hungary	100	1	1.00%
Ireland	38,790	412	1.06%
Italy	229,240	3,765	1.64%
Latvia	39	1	2.56%
Lithuania	58	0	0.00%
Luxembourg	54	1	1.85%
Malta	87	3	3.45%
Netherlands	26,600	323	1.21%

Poland	52,454	295	0.56%
Portugal	19,042	210	1.10%
Romania	24,858	409	1.65%
Slovakia	84	3	3.57%
Slovenia	20	0	0.00%
Spain	117,641	1,661	1.41%
Sweden	19,059	265	1.39%
Iceland	41	1	2.44%
Liechtenstein	14	1	7.14%
Norway	12,498	149	1.19%
Total EU	786,041	10,998	1.40%
Total EEA	798,594	11,149	1.40%
List actions per member states and languages (see example table above)	**Number of impressions of Search interventions (Covid-19 Vaccine)**	**Number of clicks of Search interventions (Covid-19 Vaccine)**	**Click Through Rate of Search interventions (Covid-19 Vaccine)**
Austria	35,708	290	0.81%
Belgium	30,470	205	0.67%
Bulgaria	2,024	38	1.88%
Croatia	149	1	0.67%
Cyprus	2,639	18	0.68%

Czech Republic	420	9	2.14%
Denmark	171	3	1.75%
Estonia	55	2	3.64%
Finland	158	1	0.63%
France	192,662	668	0.35%
Germany	259,398	1,900	0.73%
Greece	539	6	1.11%
Hungary	220	2	0.91%
Ireland	28,103	74	0.26%
Italy	160,664	626	0.39%
Latvia	35	1	2.86%
Lithuania	61	1	1.64%
Luxembourg	268	2	0.75%
Malta	66	0	0.00%
Netherlands	45,140	276	0.61%
Poland	38,819	178	0.46%
Portugal	506	1	0.20%
Romania	773	9	1.16%
Slovakia	95	0	0.00%
Slovenia	35	0	0.00%

Spain	70,179	194	0.28%
Sweden	527	5	0.95%
Iceland	15	0	0.00%
Liechtenstein	22	0	0.00%
Norway	164	2	1.22%
Total EU	869,884	4,510	0.52%
Total EEA	870,085	4,512	0.52%
List actions per member states and languages (see example table above)	**Number of impressions of Search interventions (Holocaust Denial)**	**Number of clicks of Search interventions (Holocaust Denial)**	**Click Through Rate of Search interventions (Holocaust Denial)**
Austria	48,390	6,650	13.74%
Belgium	60,452	8,132	13.45%
Bulgaria	22,821	2,885	12.64%
Croatia	363	49	13.50%
Cyprus	3,355	511	15.23%
Czech Republic	40,565	5,793	14.28%
Denmark	27,006	3,094	11.46%
Estonia	106	14	13.21%
Finland	46,272	3,495	7.55%
France	273,265	32,337	11.83%

Germany	415,587	55,986	13.47%
Greece	45,957	4,004	8.71%
Hungary	47,773	4,863	10.18%
Ireland	50,803	4,438	8.74%
Italy	158,682	16,054	10.12%
Latvia	75	7	9.33%
Lithuania	136	17	12.50%
Luxembourg	116	8	6.90%
Malta	118	11	9.32%
Netherlands	149,472	14,856	9.94%
Poland	83,102	243	0.29%
Portugal	41,015	2,854	6.96%
Romania	65,238	7,044	10.80%
Slovakia	281	25	8.90%
Slovenia	77	4	5.19%
Spain	202,141	25,280	12.51%
Sweden	76,796	9,004	11.72%
Iceland	3,342	422	12.63%
Liechtenstein	33	5	15.15%
Norway	46,042	5,008	10.88%

List actions per member states and languages (see example table above)	Number of impressions of Search interventions (Monkey Pox)	Number of clicks of Search interventions (Monkey Pox)	Click Through Rate of Search interventions (Monkey Pox)
Total EU	1,859,964	207,658	11.16%
Total EEA	1,909,381	213,093	11.16%
Austria	654	2	0.31%
Belgium	766	0	0.00%
Bulgaria	255	0	0.00%
Croatia	147	0	0.00%
Cyprus	48	0	0.00%
Czech Republic	226	1	0.44%
Denmark	574	0	0.00%
Estonia	91	1	1.10%
Finland	710	2	0.28%
France	3,859	14	0.36%
Germany	6,827	8	0.12%
Greece	634	3	0.47%
Hungary	436	0	0.00%
Ireland	852	20	2.35%
Italy	2,246	7	0.31%

Latvia	105	1	0.95%
Lithuania	216	0	0.00%
Luxembourg	35	0	0.00%
Malta	46	0	0.00%
Netherlands	1,268	3	0.24%
Poland	951	1	0.11%
Portugal	1,594	1	0.06%
Romania	2,210	6	0.27%
Slovakia	142	0	0.00%
Slovenia	83	0	0.00%
Spain	4,666	4	0.09%
Sweden	1,618	1	0.06%
Iceland	70	0	0.00%
Liechtenstein	2	0	0.00%
Norway	691	0	0.00%
Total EU	31,259	75	0.23%
Total EEA	32,022	75	0.23%
List actions per member states and languages (see example table above)	**Number of impressions of Public service announcements (Covid)**	**Number of impressions of Public service announcements (Covid 19**	**Number of impressions of Public service announcements (Holocaust Denial)**	**Number of impressions of Public service**	**Number of impressions of the safety centre page on Covid**	**Number of impressions of the safety centre page on election integrity**

		Vaccine)		announcements (Monkey Pox)		
Austria	21,993,504	1,899,751	1,170	186	11,415	510
Belgium	13,783,342	3,220,160	2,227	283	13,643	689
Bulgaria	5,112,700	599,191	562	54	12,581	999
Croatia	3,749,925	776,535	560	45	3,098	118

Cyprus	2,087,235	426,885	159	15	1,081	87
Czech Republic	6,138,525	814,921	1,562	71	2,955	226
Denmark	6,312,081	1,098,164	1,624	164	6,042	205
Estonia	1,470,707	269,449	274	11	6,490	81
Finland	7,207,338	1,381,480	1,884	147	6,986	315

France	76,737,627	29,609,328	9,477	1,093	50,022	4,635
Germany	244,810,247	17,276,470	8,630	1,877	85,491	6,067
Greece	18,648,176	2,737,857	1,529	222	7,846	702
Hungary	8,170,992	1,674,080	2,191	90	9,732	322
Ireland	10,852,789	3,725,915	1,748	155	3,140	280

Italy		100,949,203	34,503,740	3,880	522	37,610	2,347
Latvia		2,946,748	381,950	313	18	14,909	123
Lithuania		2,993,836	445,064	480	37	10,201	118
Luxembourg		1,277,896	237,442	92	23	661	63
Malta		895,625	243,943	62	10	3,395	43

Netherlands	26,638,197	4,606,618	3,625	432	36,018	1,293
Poland	38,118,818	4,289,275	5,100	269	13,792	734
Portugal	11,751,407	954,003	1,016	159	7,753	284
Romania	26,099,021	15,112,702	1,922	252	26,907	1,530
Slovakia	3,438,647	536,956	600	30	4,314	154

Slovenia	2,253,519	507,134	284	18	3,163	65
Spain	60,456,138	8,167,193	2,199	801	39,149	3,243
Sweden	14,457,314	3,003,755	3,766	410	15,720	823
Iceland	538,438	122,985	312	22	830	26
Liechtenstein	42,060	4,111	4	0	200	0

Norway	8,542,126	1,649,928	2,230	171	9,739	249
Total EU	719,351,557	138,499,961	56,936	7,394	434,114	26,056
Total EEA	728,474,181	140,276,985	59,482	7,587	444,883	26,331
Measure 17.2						
QRE 17.2.1	In order to raise awareness among our users of specific topics and empower them, we run a variety of on and off-platform media literacy campaigns. Our approach may differ depending on the topic. We localise certain campaigns (e.g., for elections) in that we collaborate with national partners and use language that the local audience can best connect with. For other campaigns such as the War on Ukraine, our emphasis was on connecting users to scalability, safety and helpful resources. We have ambitious plans to develop and roll out new media literacy campaigns in 2023, including with an initial focus on a campaign related to the War on Ukraine in Eastern European countries, which will aim to improve the digital literacy of our users. We are also continuing to work with partners such as our fact-checkers to identify specific disinformation trends and develop tailored, localised media literacy campaigns to tackle those trends.					

Below are examples of the campaigns we have most recently run in-app which have leveraged a number of the intervention tools (e.g. search interventions and video notice tags) we have outlined in our response to QRE 17.1.1.

(I) Raising awareness of disinformation DigitalMente (Italy). In collaboration with Unione Nazionale Consumatori (the Italian consumer association), in the 2021/2022 academic year, we implemented an

educational program which aimed to improve the digital skills and awareness of students aged 11-19. As part of the programme, we carefully designed and developed downloadable e-learning materials with input from industry experts (see our response to QRE 17.3.1) and are proud that the campaign was participated in by approximately 1,400 teachers and reached over 58,000 students.

(II) Promoting election integrity. As well as the elections integrity page within TikTok's Safety Center to bring visibility to how we're protecting our platform from harmful misinformation, and foreign interference, we've launched integrity campaigns in advance of several European major elections last year:

- **2022 Italian general election.** From 26 August 2022, we rolled out a campaign a month in advance of the Italian general election. The in-app Election Center we developed connected users to authoritative information sources, such as the Ministry of the Interior website as well as to our fact checkers, e.g., Facta.news. The Election Center page received over 918,000 views.

- **2022 Danish election.** From 6 October 2022, we rolled out an in-app campaign a month in advance of the Danish election designed to educate users on fact-checking and connect them to authoritative information sources. We implemented the campaign in partnership with Danish fact-checker Tjekdet and linked users to their website. The campaign page received over 13,554 views.

- **2022 French presidential and parliamentary elections.** From 17 February 2022, we launched an in-app Election Centre to provide users up-to-date information about the 2022 French presidential and legislative elections. The centre contained a section about fighting-fake news, which linked to media literacy association Génération Numérique. The election page on the presidential election received around 1.3 million views and the page on the legislative election was viewed more than 220,000 times.

(III) War in Ukraine. We rolled out an in-app digital literacy campaign specific to information about the crisis, which encouraged users to verify and evaluate information and provided links to United Nations webpages to facilitate user donations. We implemented localised versions of the campaign in 18 European countries, however we do not have engagement data available. We are working on resolving this for future campaigns.

(IV) Covid-19. We implemented a comprehensive campaign across a number of jurisdictions to address disinformation related to Covid-19. As part of this we are proud to have created #TeamHalo, which continues as a means to update users on vaccines. Our fact-checking partners continue to be heavily involved in the campaign and our dedicated notice tags and search intervention tools now direct users to authoritative, localised information from expert organisations (such as local public health sites or, where local health sites are not available, the WHO).

SLI 17.2.1 - actions enforcing policies above	N/A				
	N/A	N/A	N/A	N/A	N/A

Member States					
List actions per member states and languages (see example table above)					
Measure 17.3					
QRE 17.3.1	As outlined in our Safety Partners page within TikTok's Safety Center, we partner with experts on many areas including on media literacy campaigns including to collaborate on the development of the campaign and to direct users to their authoritative resources. Our experts include fact-checkers, academics, NGOs and government entities. Specific examples of partnerships within the campaigns and projects set out in QRE 17.2.1 are:				

(I) Raising awareness of disinformation

DigitalMente (Italy). We implemented the campaign in partnership with Unione Nazionale Consumatori (the Italian consumer association). The e-learning materials were designed with input from privacy experts and behavioural psychologists who specialise in the area of online safety, such as psychologist Maura Manca, who has particular expertise in the area of teens and digital wellbeing. These materials were made available on the Italian digital education platform Educazione Digitale, which is endorsed by the Italian Ministry of Education.

(II) Promoting election integrity. We partner with various media organisations and fact-checkers in the context of our election campaigns. For more detail about the input our fact-checking partners provide please refer to QRE 30.1.3.
- For the Italian and Danish elections, we partnered with Facta.News and TjekDet respectively.
- For the French presidential and parliamentary elections, we partnered with Génération Numérique, an association endorsed by the National Ministry of Education, focusing on educating students and parents about issues that arise in the digital sphere. Génération Numérique works in close collaboration with government and regulatory entities including the CNIL and the Ministries of Education, Culture, Children and the Interior. We also worked with our fact-checking partner Agence France Press in relation to the assessment of election related content.

(III) Covid-19. We again partnered with fact checkers throughout the EU in relation to Covid-19 misinformation, including AFP, Facta, Logically, Lead Stories, Newtral, Science Feedback, Teyit and DPA. We have also continued our partnership with the WHO Tech Taskforce and Meedan, both of whom provide us with information about Covid-19 trends including new variants, vaccine side effects on adults and children, and information about new treatments (anti-virals, etc). Finally, we've created #TeamHalo, which allows top scientists and clinicians to share their vaccine stories and authoritative vaccine information with users.

V. Empowering Users	
Commitment 18	
Relevant Signatories commit to minimise the risks of viral propagation of Disinformation by adopting safe design practices as they develop their systems, policies, and features.	
In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	Yes
If yes, list these implementation measures here [short bullet points].	Continual improvement in detecting and removing misinformation by investing in: ● machine learning models and increasing capacity to iterate on these models rapidly given the fast changing nature of misinformation; ● improved detection of known misleading audio and imagery to reduce manipulated content; and ● a proactive detection program with our fact-checkers who flag new and evolving claims they're seeing across the internet. This allows us to look for these claims on our platform and remove violations. These continued investments have brought gains to our proactive detection and enforcement of these policies.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	● Driving forward our ambitious plans to further expand our fact-checking network, by adding new fact-checking partners and increasing the number of languages our current fact-checking partners provide support in. Over the next year, we will be focusing on setting up fact-checking partnerships in Portugal, Denmark, Greece and Belgium as well as other EU countries. As part of these plans, we will be expanding our misinformation moderation teams. ● Expanding the application of our state-controlled media label and the capabilities and use-cases of our unverified content label. ● Investing further in our detection system and testing new machine learning models covering EU languages to tackle harmful misinformation at scale. ● Continue investing in training and development for our Trust and Safety team with the help of our fact-checking partners and external experts in this field.

	• Launching a pilot phase as part of our efforts to develop an API for researchers, as we are committed to engaging with the research community and ensuring that our tools meet their needs going forwards.
Measure 18.1	
QRE 18.1.1	Our most important commitment at TikTok is to create a welcoming environment that keeps our community safe and this means minimising the risks of our community being exposed to the viral spread of harmful misinformation on our platform. Our approach to safety is systemic and spans policies, product, practices, and partners.
	First, safety is built into our recommendations. We have a number of safeguards in place to ensure the For You feed (as the primary access point for content on the platform) has safety built-in. Key to reducing the prominence of harmful misinformation on our platform is removing harmful misinformation that violates our CGs:
	(I) Removal of content or accounts. Our I&A policies make clear that we do not allow activities that may undermine the integrity of our platform or the authenticity of our users. We remove content or accounts that involve misleading information that causes significant harm or, in certain circumstances, reduce the prominence of content. **QRE 14.1.1** and **QRE 18.2.1** outline in detail our policies and the enforcement action we take.
	(II) Proactive moderation. We place considerable emphasis on proactive content moderation, which means that our Trust and Safety teams work to detect and remove harmful misinformation *before* it is reported to us by users or third parties. The results of our proactive moderation efforts can be seen in the data reported in our [Transparency Reports](#) under "proactive removals". Proactive moderation operates in a number of ways at TikTok:
	• **Automated review.** Once uploaded to the platform, video content goes through a technology-based automated review process designed to flag content that may violate our I&A policies. This happens in real time given the volume of video content posted to the platform on a continuous basis. If the video is flagged during the automated review and it is determined that it may contain harmful misinformation, it will, in turn, be flagged for human review by TikTok's specialised misinformation moderators. Whilst the video is being reviewed by our misinformation moderators or our fact-checking partners, it will not be eligible to be served to the For You feed. We also use models to review text and audio content to detect potentially harmful misinformation (such as via keyword lists, and using natural language processing techniques). Video content also undergoes review by deduplication models which compare newly uploaded video content to match hashed content which has previously been found to violate our CGs.

- **Targeted sweeps:** TikTok's Trust and Safety team from time to time also undertake targeted sweeps of certain types of violative content including potential harmful misinformation, where they have identified specific risks. For example, this may involve a targeted sweep focused on use of a specific keyword, URL or hashtag identified as being associated with high risk content. Such content may then undergo manual review by our moderators in order to detect and mitigate safety risks.
- **Fact-checkers and experts.** We work with our fact checkers and outside experts to understand what kind of misinformation trends are occurring on other platforms. By being alert to content trends outside of our platform, we can improve automated review, carry out targeted sweeps and issue guidance to our moderation teams to help them more easily spot and take swift action on violating content.

(III) Reporting content or accounts. Anyone can report potential harmful misinformation or accounts on our platform, which may trigger manual review by our moderators and if appropriate, enforcement action. Our fact-checking partners can also report potential harmful misinformation to us directly.

(IV) Safety in our recommendations. In addition to removing content that clearly violates our CGs, our safety team takes additional precautions to review videos as they rise in popularity. Such review is undertaken given the extent of the content's dissemination and the increase in potential harm if the content is found to be in breach of our I&A policies.

a. For content that does not violate our CGs but may negatively impact the authenticity of the platform, we reduce its prominence and / or label it. As mentioned, sometimes the authenticity of content cannot be confirmed, especially during unfolding events. We may limit the video's distribution by not showing it in anyone's For You feed and we may label it as "unverified content". We also label accounts and content of state affiliated media entities to empower users to consider the sources of information. Once we identify a content trend e.g., Covid-19, we can deploy a number of in-app tools, such as applying video notice tags, in order to improve the media literacy of our users.
b. We may also increase the prominence of authoritative information. Again once we identify a content trend, we can deploy a number of in-app tools including a banner which will redirect users to authoritative sources.

Second, safety by design is built into our platform and all of the features. Within our Trust and Safety team, we have individuals in product and policy and subject matter experts dedicated to the topic of integrity and authenticity. In developing a new feature or policy, these teams our Trust and Safety team work closely with external partners to incorporate their expertise and ensure we are reflecting industry best practice. For example (as outlined in **QRE 21.3.1**), we collaborated with Irrational Labs to develop and implement specialised prompts to help users consider before sharing unverified content.

	Please refer to the response to QRE 19.1.1 where we have outlined the main parameters of our recommender system.			
QRE 18.1.3	Please refer to the response to QRE 18.1.1.			
SLI 18.1.1 - actions proving effectiveness of measures and policies	**Methodology of data measurement:** The share cancel rate (%) following the unverified content label share warning pop-up indicates the percentage of users who do not share a video after seeing the label pop up. This metric is based on the active location of the users that engaged with these tools.			
	N/A	N/A	N/A	Other relevant metrics
Member States				
List actions per member states and languages (see example table above)				**Share cancel rate (%) following the unverified content label share warning pop-up (users who do not share the video after seeing the pop up)**
Austria				28.57%
Belgium				27.72%
Bulgaria				21.05%
Croatia				23.08%
Cyprus				11.11%
Czech Republic				30.77%
Denmark				44.83%
Estonia				30.77%

Finland				32.43%
France				32.43%
Germany				25.87%
Greece				25.00%
Hungary				41.18%
Ireland				22.10%
Italy				36.63%
Latvia				24.14%
Lithuania				33.33%
Luxembourg				36.36%
Malta				0.00%
Netherlands				27.31%
Poland				29.40%
Portugal				23.64%
Romania				23.98%
Slovakia				50.00%
Slovenia				33.33%
Spain				38.78%
Sweden				24.73%
Iceland				50.00%

Liechtenstein				0.00%
Norway				23.08%
Total EU				28.84%
Total EEA				28.39%
Measure 18.2				

QRE 18.2.1	As stated, to help keep our platform welcoming and authentic for everyone, we take seriously ensuring the content on our platform is free from harmful misinformation. Our I&A policies within our CGs and Terms of Service are the first line of defence in combating harmful misinformation and (as outlined in more detail in QRE 14.1.1) deceptive behaviours on our platform. These guidelines make clear to our users what content we remove or make ineligible for the For You feed when they pose a risk of harm to our users and their communities. These CGs are also explained to our users in a series of user-friendly, colourful [videos](#) in our Safety Center. As is made clear in our CGs, we do not allow misinformation that causes significant harm to individuals, our community, or the larger public regardless of intent. Specifically, our policies do not allow: • Misinformation that incites hate or prejudice. • Misinformation related to emergencies that induces panic. • Medical misinformation that can cause harm to physical health. • Content that misleads about elections or other civic processes. • Conspiracy theories that may cause hate or prejudice, attack a specific person or a protected group, make a violent call to action, deny a violent or tragic event occurred, or have been linked to violence. As outlined in the QRE 14, we also remove accounts that seek to mislead people or use TikTok to deceptively sway public opinion. These activities range from inauthentic or fake account creation, to more sophisticated efforts to undermine public trust. As stated, we have policy experts within our Trust and Safety team dedicated to the topic of integrity and authenticity. They continually keep these policies under review and collaborate with external partners and experts when understanding whether updates or new policies are required. **Enforcing our policies.** We remove content – including video, audio, livestream, images, comments, links, or other text – that violates our I&A policies. Individuals are notified of our decisions and can appeal them if they believe no violation has occurred. We also make clear in our CGs that we will temporarily or

permanently ban accounts and/or users that are involved in severe or repeated on-platform violations, this includes breaches of our I&A policies.

Similar to how we enforce other content issues under the CGs, we enforce our I&A policies through a mix of technology and human moderation. To do this effectively at scale, we continue to invest in our automated review process as well as in people and training. At TikTok we place a considerable emphasis on proactive content moderation. This means our teams work to detect and remove harmful material before it is reported to us.

However, misinformation is different from other content issues. Context and fact-checking are critical to consistently and accurately enforcing our misinformation policies. So while we use machine learning models to help detect potential misinformation, ultimately our approach today is having our moderation team assess, confirm, and remove misinformation violations. We have specialised misinformation moderators who have enhanced training, expertise, and tools to take action on harmful misinformation. This includes direct access to our fact-checking partners who help assess the accuracy of content.

We strive to maintain a balance between freedom of expression and protecting our users and the wider public from harmful content. Our approach to combating harmful misinformation, as stated in our CGs, is to remove content that is both false and can cause harm to individuals or the wider public. This does not include simply inaccurate information which does not pose a risk of harm. Additionally, in cases where fact checks are inconclusive, especially during unfolding events, the video may not be removed and may instead become ineligible for recommendation into anyone's For You feed to limit the spread of potentially misleading information and labelled with the "unverified content" label. We are working to improve the accuracy of the data metrics in respect of not recommend moderation actions and will include these in future reports once that work is complete.

Note that in relation to the metrics we have shared at SLI 18.2.1 below, of all the views that occurred in Q4 2022, fewer than 1 in 10,000 views occurred on content identified and removed for violating our policies around harmful misinformation. We appreciate that this does not capture views of content that was not identified and we will be exploring new ways to measure exposure to this type of content in the future.

SLI 18.2.1 - actions taken in response to policy violations	**Methodology of data measurement:**		
	We have based the following numbers on the country in which the video was posted: videos removed because of violations of our harmful misinformation policies.		
	The number of views of videos removed because of violation of each of the harmful misinformation policies is based on the active location of the user.		
	Total no of violations	Metric 1: indicating the impact of the action taken	N/A	N/A

Member States				
List actions per member states and languages (see example table above)	**Number of videos removed because of violation of harmful misinformation policy**	**Number of views of videos removed because of violation of harmful misinformation policy**		
Austria	174	2,982,119		
Belgium	4	0		
Bulgaria	38	760,281		
Croatia	7	71,425		
Cyprus	7	279,487		
Czech Republic	220	5,587,618		
Denmark	29	1,031,958		
Estonia	1	137,428		
Finland	73	758,927		
France	9,442	7,129,562		
Germany	5,855	72,249,825		
Greece	207	1,943,434		
Hungary	202	5,519,983		
Ireland	485	8,466,010		
Italy	10,491	16,293,085		
Latvia	1	35,074		

Lithuania	3	1,133,779		
Luxembourg	0	10,007		
Malta	0	0		
Netherlands	401	2,667,892		
Poland	2,006	19,713,692		
Portugal	123	7,817,169		
Romania	912	8,259,483		
Slovakia	36	2,674,590		
Slovenia	4	61,423		
Spain	4,918	11,222,311		
Sweden	868	6,581,188		
Iceland	11	1,271,011		
Liechtenstein	0	0		
Norway	59	1,431,131		
Total EU	36,507	183,387,750		
Total EEA	36,577	186,089,892		
Measure 18.3				
QRE 18.3.1	We have partnered with third party experts and researchers in relation to the creation of certain of our warning and labelling systems, which are designed to reduce the spread of disinformation. Most notably, our close work with behavioural psychologists, Irrational Labs, led to the development of the following features (both of which are explained in further detail at QRE 17.1.1):			

	• [specialised prompts for unverified content](#) which, show viewers of unverified content a banner cautioning them that the video has been flagged for unverified content; and • our [state controlled media label,](#) which brings transparency to our community in relation to state controlled media entities and raises awareness among users to encourage users to consider the reliability of the source. Further information regarding the specific ways in which we worked with Irrational Labs in relation to the design of these features is set out at QRE 21.3.1.

V. Empowering Users
Commitment 19 Relevant Signatories using recommender systems commit to make them transparent to the recipients regarding the main criteria and parameters used for prioritising or deprioritising information, and provide options to users about recommender systems, and make available information on those options.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	Yes
If yes, list these implementation measures here [short bullet points].	• Further empowered users with ways to customise their viewing experience, by rolling out a tool people can use to automatically filter out videos with words or hashtags they don't want to see in their For You or Following feeds. • Introduced the "Why this video" feature, which builds into our app transparency in relation to how our recommender system works and empowers our users to better understand why a particular video has been recommended to them. • Revamped the design of the Safety Center website to increase the ease of use, by making it more colourful and using more images and graphics and creating a series of longer length videos to bolster our efforts to communicate with users, specifically on the For You feed and our recommendation system.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes

If yes, which further implementation measures do you plan to put in place in the next 6 months?	- Further developing a new system we have designed to organise content based on thematic maturity. The system aims to enable users to avoid certain categories of content based on informed consent or their comfort with certain types of content. The version we have already rolled out helps prevent content with overtly mature themes from reaching audiences between ages 13-17. In the coming year, we plan to continue developing the system and adding functionality to support our content classification efforts. - Adding new functionality to provide detailed content filtering options for our entire community so they can enjoy more of what they love.
Measure 19.1	
QRE 19.1.1	The "For You" feed is the interface users first see when they open TikTok. It's central to the TikTok experience and where most of our users spend their time exploring the platform. We make clear to users in our Terms of Service and CGs (and also provide more context in a newsroom post and our Safety Center) that each account holder's "For You" feed is based on a personalised recommendation system. The "For You" feed is curated to each user. As stated above, safety is built into our recommendations. As well as removing harmful misinformation content that violates our CGs, we take steps to avoid recommending certain categories of content that may not be appropriate for a general audience including general conspiracy theories. **Main parameters.** The system recommends content by ranking videos based on a combination of factors including: - user selections (i.e. interest categories indicated in the registration process or selecting "not interested" on content); - user interactions (i.e. videos users like/share, accounts users follow, comments users post and content they create); - video information (i.e. captions/sounds and hashtags used); and - device and account settings (i.e. language preferences, country settings and device types). All these factors are weighted based on their value to a user and the recommendation system then ranks them to determine the likelihood of a user's interest in a particular category of content. If a user watches a longer video from beginning to end, this would be considered a strong indicator of interest and receive a greater weight than a weak indicator such as device and account settings because users don't actively express these preferences. Examples of factors which do not impact the recommendation system include user follower count and whether users have had high-performing videos previously. Building on our existing recommender system transparency we recently introduced the "Why this video" feature, which allows users to see with any particular video that appears in their For You feed factors that

	influenced why it appeared in their feed. This feature provides added transparency in relation to how our ranking system works and empowers our users to better understand why a particular video has been recommended to them. The feature essentially explains to users how past interactions on the platform have impacted the video they have been recommended. For further information, see our newsroom post.
	Diversifying recommendations
	User preferences. Together with the safeguards we build into our platform by design, we also empower our users to customise their experience to their preferences and comfort.
	- Users can click on any video and select "not interested" to automatically skip future videos from that same creator or using the same audio. - Since July 2022, users have been able to automatically filter out videos with specific words or hashtags associated with content that they don't want to see from their For You or Following feeds. - Users can enable "Restricted Mode" within their account settings in order to limit the appearance of content that may not be appropriate for all audiences. - We have been developing a new system which is designed to organise content based on thematic maturity, which aims to enable users to avoid certain categories of content based on informed consent or their comfort with certain types of content. The version already rolled out is helping to prevent content with overtly mature themes from reaching audiences between ages 13-17. In the coming year, we plan to continue developing the system and adding new functionality to support our content classification efforts.
	In the context of TikTok's obligations under the DSA, we are also actively exploring mechanisms for giving users options to modify or influence what they see on our platform.
	We recently overhauled the design of the Safety Center website to make it more colourful and to use more images and graphics and created a series of longer length videos to bolster our efforts to communicate with users via a variety of media formats. Thus, as well as providing this information to users in our Terms of Service and CGs, we have created a series of informative videos specifically on the For You feed and our recommendation system. Users can also browse through the top questions we get asked on our recommendation systems and our answers.
	As well as making clear to users how the recommender system works, we have established our Transparency and Accountability Centers, where there's an opportunity to get even more detail on how our recommendation systems work.
Measure 19.2	

SLI 19.2.1 – user settings	Methodology of data measurement: The number of times a user clicks "not interested" in relation to the For You feed and the number of users who have filtered hashtags or a keyword to set preferences for For You feed are both based on the active location of the users that engaged with these tools.			
	No of times users actively engaged with these settings	No of times users actively engaged with these settings		
Member States				
List actions per member states and languages (see example table above)	**Number of users that filtered hashtags**	**Number of users that clicked on "not interested"**		
Austria	14,694	749,673		
Belgium	18,838	1,112,001		
Bulgaria	8,902	591,195		
Croatia	5,978	287,711		
Cyprus	1,610	133,541		
Czech Republic	13,350	724,251		
Denmark	11,047	486,811		
Estonia	3,568	126,067		
Finland	12,453	568,897		
France	90,402	6,868,298		
Germany	145,796	7,535,885		
Greece	15,850	990,076		

Hungary	13,405	831,194		
Ireland	15,977	669,785		
Italy	78,802	5,522,119		
Latvia	4,527	209,459		
Lithuania	6,050	248,454		
Luxembourg	1,417	66,572		
Malta	819	57,838		
Netherlands	40,615	2,048,600		
Poland	59,839	3,027,480		
Portugal	16,718	930,575		
Romania	23,367	2,303,835		
Slovakia	5,445	282,006		
Slovenia	2,891	134,808		
Spain	76,932	5,193,006		
Sweden	25,917	1,194,059		
Iceland	1,409	44,473		
Liechtenstein	46	1,737		
Norway	15,489	593,160		
Total EU	715,209	42,894,196		
Total EEA	732,153	43,533,566		

V. Empowering Users	
Commitment 20	
Relevant Signatories commit to empower users with tools to assess the provenance and edit history or authenticity or accuracy of digital content.	
In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	TikTok did not subscribe to this commitment. TikTok considers that it would be imprudent to commit to this measure at a time when the underlying technology remains unproven and the standards to be complied with are not yet finalised.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No
If yes, which further implementation measures do you plan to put in place in the next 6 months?	N/A
Measure 20.1	
QRE 20.1.1	**Not committed**. TikTok considers that it would be imprudent to commit to this measure at a time when the underlying technology remains unproven and the standards to be complied with are not yet finalised. Once the relevant technology is proven and the standards are agreed and identifiable, TikTok will assess whether such an approach would be beneficial, taking into account existing measures in place.
Measure 20.2	
QRE 20.2.1	**Not committed**. Similarly to Measure 20.1, TikTok considers that it would be imprudent to commit to this measure at this time. TikTok is, however, open to reassessing in the future whether such an approach would be beneficial.

V. Empowering Users

Commitment 21
Relevant Signatories commit to strengthen their efforts to better equip users to identify Disinformation. In particular, in order to enable users to navigate services in an informed way, Relevant Signatories commit to facilitate, across all Member States languages in which their services are provided, user access to tools for assessing the factual accuracy of sources through fact-checks from fact-checking organisations that have flagged potential Disinformation, as well as warning labels from other authoritative sources.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	Yes
If yes, list these implementation measures here [short bullet points].	Secured new fact-checking partnerships in Sweden, Hungary, Poland and Romania and expanded our misinformation moderation teams for these countries.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	- Consulting with external researchers in relation to the development of a feature which would facilitate engagement with users who have interacted with harmful misinformation on our platform. - Expanding the capabilities and use-cases of the unverified content label feature. - Driving forwards our ambitious plans to further expand our fact-checking network, by adding new fact-checking partners and increasing the number of languages our current fact-checking partners provide support in. Over the next year, we will be focusing on setting up fact-checking partnerships in Portugal, Denmark, Greece and Belgium as well as other EU countries.
Measure 21.1	
QRE 21.1.1	We currently have 8 IFCN accredited fact checking organisations providing coverage in Europe, in 10 of the official European languages. Our response to QRE 30.1.2 sets out the specific organisations we partner with. We ensure that our users benefit from the context and insights provided by the fact checking organisations we partner with in the following ways:

	● **Enforcement of misinformation policies.** Our fact-checking partners play a critical role in helping us enforce our misinformation policies, which aim to promote a trustworthy and authentic experience for our users. We consider context and fact-checking to be key to consistently and accurately enforcing these policies, so, while we use machine learning models to help detect potential misinformation, we have our specialised misinformation moderators assess, confirm, and take action on harmful misinformation. As part of this process, our moderators are able to provide content to our expert fact checking partners for further evaluation. Where fact checkers advise that content is false, our moderators take measures to assess and remove it from our platform. Our response to QRE 31.1.1 provides further insight into the way in which fact-checkers are involved in this process.
	
	● **Unverified content labelling.** Our fact-checkers support certain of the in-app tools we have designed in order to bring users additional context on certain content or provide access to authoritative information. As mentioned above, we partner with fact checkers to assess the accuracy of content. Sometimes, our fact-checkers determine that content is not able to be confirmed or checks are inconclusive (especially during unfolding events or crises). Where our fact-checkers provide us with an inconclusive/ unconfirmed determination, we use our unverified content label to inform viewers [via a banner](#) that a video contains unverified content, in an effort to raise user awareness about content credibility. In these circumstances, the content creator is also notified that their video was flagged as unsubstantiated content.
	
	● **In-app tools related to specific topics:**
	    ○ **Covid-19.** We partnered with a number of fact checkers throughout the EU in order to prevent the spread of harmful misinformation related to Covid-19 on our platform, including AFP, Facta, Logically, Lead Stories, Newtral, Science Feedback, Teyit and DPA.
	
	    ○ **Election integrity.** We have launched campaigns in advance of several major elections aimed at educating the public about the voting process which encourage users to fact-check information with our fact-checking partners. The election integrity campaign we rolled out in advance of the Italian general election included setting up an in-app [Elections Center](#) to connect people who engage with election content with our fact checkers, e.g., [Facta.news](#). We launched the campaign with a [blog post](#) that introduced the Election Center and explained our content labelling and fact-checking processes.
	
	● **User awareness of our fact-checking partnerships and labels.** We have published [blog](#) posts (available in more than 25 languages) and created a hub on our Safety Center to raise users' awareness of our fact-checking programme and labels and to support the work of our fact checking partners.

SLI 21.1.1 - actions taken under measure 21.1	**Methodology of data measurement:** The share of removals under our harmful misinformation policy is relative to total removals under our CGs.

	The share of proactive removals, share of removals before any views and share of the removals within 24h are relative to total removals under our harmful misinformation policies. The share cancel rate (%) following the unverified content label share warning pop-up indicates the percentage of users who do not share a video after seeing the label pop up. This metric is based on the active location of the users that engaged with these tools.				
	Reach of labels/ fact-checkers and other authoritative sources	Other pertinent metric	Other pertinent metric	Other pertinent metric	Other pertinent metric
Member States					
List actions per member states and languages (see example table above)	**Share cancel rate (%) following the unverified content label share warning pop-up (users who do not share the video after seeing the pop up)**	**Share of removals under harmful misinformation policy**	**Share of proactive removals under misinformation policy**	**Share of removals before any views under misinformation policy**	**Share of the removals within 24 h by misinformation policy**
Austria	28.57%	0.23%	0.12%	0.09%	0.07%
Belgium	27.72%	0.00%	0.00%	0.00%	0.00%
Bulgaria	21.05%	0.03%	0.01%	0.00%	0.01%
Croatia	23.08%	0.02%	0.03%	0.01%	0.02%
Cyprus	11.11%	0.05%	0.01%	0.01%	0.02%
Czech Republic	30.77%	0.20%	0.01%	0.01%	0.12%
Denmark	44.83%	0.05%	0.00%	0.00%	0.03%

Estonia	30.77%	0.00%	0.00%	0.00%	0.01%
Finland	32.43%	0.10%	0.02%	0.02%	0.05%
France	32.43%	1.41%	1.53%	2.07%	1.44%
Germany	25.87%	0.72%	0.62%	0.60%	0.11%
Greece	25.00%	0.15%	0.06%	0.00%	0.09%
Hungary	41.18%	0.17%	0.05%	0.03%	0.10%
Ireland	22.10%	0.38%	0.40%	0.41%	0.35%
Italy	36.63%	1.57%	1.61%	1.80%	1.70%
Latvia	24.14%	0.00%	0.00%	0.00%	0.00%
Lithuania	33.33%	0.01%	0.00%	0.00%	0.00%
Luxembourg	36.36%	0.00%	0.00%	0.00%	0.00%
Malta	0.00%	0.00%	0.00%	0.00%	0.00%
Netherlands	27.31%	0.13%	0.10%	0.10%	0.10%
Poland	29.40%	0.30%	0.27%	0.30%	0.25%
Portugal	23.64%	0.12%	0.03%	0.01%	0.10%
Romania	23.98%	0.25%	0.25%	0.24%	0.19%
Slovakia	50.00%	0.08%	0.01%	0.00%	0.02%
Slovenia	33.33%	0.02%	0.02%	0.01%	0.03%
Spain	38.78%	0.80%	0.81%	1.06%	0.93%
Sweden	24.73%	0.45%	0.49%	0.54%	0.41%

Iceland	50.00%	0.21%	0.10%	0.02%	0.14%
Liechtenstein	0.00%	0.00%	0.00%	0.00%	0.00%
Norway	23.08%	0.08%	0.02%	0.01%	0.06%
Total EU	28.84%	0.27%	0.24%	0.27%	0.23%
Total EEA	28.39%	0.25%	0.22%	0.24%	0.21%

SLI 21.1.2 - actions taken under measure 21.1	**Methodology of data measurement:** The number of videos tagged with the unverified content label is based on the country in which the video was posted.		
	N/A	Nr of labels applied to content, such as on the basis of such articles	Meaningful metrics such as the impact of 21.1. measures on user interactions with, or user re-shares of, content fact-checked as false or misleading
Member States			
List actions per member states and languages (see example table above)		**Number of videos tagged with the unverified content label**	**Share cancel rate (%) following the unverified content label share warning pop-up (users who do not share the video after seeing the pop up)**
Austria		1,036	28.57%
Belgium		1,288	27.72%
Bulgaria		405	21.05%
Croatia		247	23.08%

Cyprus		217	11.11%
Czech Republic		578	30.77%
Denmark		389	44.83%
Estonia		341	30.77%
Finland		540	32.43%
France		3,756	32.43%
Germany		3,820	25.87%
Greece		494	25.00%
Hungary		375	41.18%
Ireland		1,096	22.10%
Italy		2,132	36.63%
Latvia		465	24.14%
Lithuania		469	33.33%
Luxembourg		219	36.36%
Malta		183	0.00%
Netherlands		1,543	27.31%
Poland		1,767	29.40%
Portugal		600	23.64%
Romania		699	23.98%
Slovakia		286	50.00%

Slovenia		196	33.33%
Spain		2,456	38.78%
Sweden		769	24.73%
Iceland		115	50.00%
Liechtenstein		19	0.00%
Norway		584	23.08%
Total EU		26,366	28.84%
Total EEA		27,084	28.39%
QRE 21.2.1	We have been progressing this research project internally with the relevant stakeholders and are finalising plans which should enable us to commence work on this feature, in consultation with researchers, in early 2023.		
Measure 21.3			
QRE 21.3.1	As set out within our response to QRE 17.1.1, we apply our unverified content and state-controlled media labels to certain content in order to empower our community by providing them with an additional layer of context. We ensure these labels are developed and deployed in accordance with scientific evidence by partnering with fact-checkers and working with external experts, including scientists, in the following ways:		

- **Unverified content label.** As mentioned above, we partner with a number of IFCN registered fact checkers in Europe, who assist with assessing the accuracy of certain content on our platform. Where our fact-checking partners determine that a video is not able to be confirmed or their fact-checks are inconclusive (which is sometimes the case, particularly during unfolding events or crises), we may apply our unverified content label to the video.

We also partnered with behavioural scientists, Irrational Labs, in relation to the roll out of the specialised prompts which encourage users to consider content which has been labelled as unverified, before sharing it. The prompts, which are explained in further detail at QRE 17.1.1, were designed and tested with Irrational Labs. On testing the prompts, Irrational Labs found that viewers decreased the rate at which they shared videos by 24%, while likes on such unsubstantiated content also decreased by 7%.

	- **State-controlled media label**. In the last year, we have started applying our state-controlled media label to accounts or content belonging to media organisations whose editorial decision making is directly or indirectly controlled by a government. To inform our definition of state-affiliated media and approach to making such designations, we engaged with over 50 experts across 20 countries and will continue to work with these experts to inform our global approach and expansion of the policy. As mentioned above, one of the experts we worked closely with was Irrational Labs. In collaboration with the behavioural scientists, we considered the ways in which we could present the label to our users. As part of this, we tested various copy options across English, Spanish, and Arabic via quantitative surveys and qualitative panels, and found that "[country] state-controlled media" was the option most preferred by users while being the most accurate representation of the relevant media entities' relationship to their respective governments. We note that we are also continuously taking user feedback into consideration in order to identify new topics and consider which tools may be best suited to raising awareness around that topic and combating harmful misinformation.

V. Empowering Users
Commitment 22
Relevant Signatories commit to provide users with tools to help them make more informed decisions when they encounter online information that may be false or misleading, and to facilitate user access to tools and information to assess the trustworthiness of information sources, such as indicators of trustworthiness for informed online navigation, particularly relating to societal issues or debates of general interest.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	TikTok did not subscribe to this commitment.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes

If yes, which further implementation measures do you plan to put in place in the next 6 months?	Aiming to build products and processes to more quickly launch user education interventions when a new crisis emerges.
Measure 22.1	
QRE 22.1.1	**Not committed.** TikTok considers that the implementation of the other commitments under the Code (including, but not limited to, the other commitments relating to empowering users and fact-checking) provide a comprehensive approach to tackling disinformation, including facilitating users in making more informed decisions when they encounter online information that may be false or misleading. This position is also reflected in the Commission's guidance. As such, TikTok does not consider that committing to this measure will materially add to the other measures being adopted under the Code.

TikTok does, however, keep an open mind. As such, it is prepared to explore such solutions with the relevant providers in due course and to reassess whether such an approach would be beneficial, taking into account existing measures in place. |
SLI 22.1.1 - actions enforcing policies above	N/A
	N/A
Member States	
List actions per member states and languages (see example table above)	
Measure 22.2	
QRE 22.2.1	**Not committed.** TikTok does not consider this to be a practical or implementable proposal. In any event, this measure is unnecessary as the combination of other commitments underpinning the Code (including, but not limited to, those relating to fact-checking) represent a comprehensive approach to achieving the goal of providing users with tools to make more informed decisions when they encounter online information that may be false or misleading.
Measure 22.3	
QRE 22.3.1	**Not committed.** Not relevant, linked to the above Measures.
Measure 22.4	

QRE 22.4.1	**Not committed.** Measures 22.4, 22.5 and 22.6 are not applicable as TikTok is not a provider of trustworthiness indicators.
SLI 22.4.1 - actions enforcing policies above	N/A
	N/A
Member States	
List actions per member states and languages (see example table above)	
Measure 22.5	
QRE 22.5.1	**Not committed.** Measures 22.4, 22.5 and 22.6 are not applicable as TikTok is not a provider of trustworthiness indicators.
SLI 22.5.1 - actions enforcing policies above	N/A
	N/A
Member States	
List actions per member states and languages (see example table above)	
SLI 22.5.2 - actions enforcing policies above	N/A
	N/A
Member States	
List actions per member states and languages (see example table above)	
Measure 22.6	
QRE 22.6.1	**Not committed.** Measures 22.4, 22.5 and 22.6 are not applicable as TikTok is not a provider of trustworthiness indicators.

SLI 22.6.1 - actions enforcing policies above	N/A
	N/A
Member States	
List actions per member states and languages (see example table above)	
Measure 22.7	
QRE 22.7.1	As per our response to QRE 17.1.1, we have numerous tools (video notice tags, search interventions, public service announcements, in-app information hubs and Safety Center pages) that lead users to authoritative sources available in all EU member states and in 22 EU languages. We also run localised campaigns on specific topics which deploy different engagement techniques depending on the subject matter and / or member state involved, e.g., in-person workshops, radio and newspaper campaigns.

SLI 22.7.1 - actions enforcing policies above	N/A			
	N/A	N/A	N/A	N/A
Member States				
List actions per member states and languages (see example table above)				

V. Empowering Users
Commitment 23
Relevant Signatories commit to provide users with the functionality to flag harmful false and/or misleading information that violates Signatories policies or terms of service.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No

If yes, list these implementation measures here [short bullet points].	N/A
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/ No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	We will continue to invest in resources and training for our misinformation moderators.
Measure 23.1	
QRE 23.1.1	We provide users with simple, intuitive ways to report/flag content in-app for any breach of our Terms of Service or CGs including for harmful misinformation in each EU member state and in an official language of the European Union. ● By 'long-pressing' (e.g., clicking for 3 seconds) on the video content and selecting the "Report" option. The user is then shown categories of reporting reasons from which to select (which align with the harms our CGs seek to address) including "*harmful misinformation*" and "*spam and fake engagement*". ● By selecting the "Share" button available on the right-hand side of the video content and then selecting the "Report" option. Users do not need to be logged into an account on the platform to report content, and can also report video content via the TikTok website (by clicking on the "Report" button which is prominently displayed in the upper right hand corner of each video when hovering over a video) or by means of our "Report Inappropriate content" webform which is available in our Support Centre. We are aware that harmful misinformation is not limited to video content and so users can also report a comment, a suggested search, a hashtag, a sound or an account, again specifically for harmful misinformation.
Measure 23.2	
QRE 23.2.1	**Appeals system**

The integrity of our appeals systems is predominantly ensured through the involvement of our specialised human moderators, who are able to take context into account when reaching decisions in relation to whether content complies with our CGs and is therefore safe to be shared on our platform.

Our moderators review all appeals raised in relation to removed videos, removed comments and banned accounts.

In order to reach a decision as to whether the user's content can be reinstated on our platform, our moderators review the policies set out in our CGs, and consider whether such content is in violation of those policies. To ensure consistency within this process and its overall integrity, we have sought to make our CGs as clear and comprehensive as possible and have put in place robust Quality Assurance processes (including steps such as auditing appeals and undertaking Root Cause Analyses).

Our users have the ability to share feedback with us to the extent that they don't agree with the determination of our moderators in relation to their appeal. They can do so by using the function which allows them to "report a problem". We are continuously taking user feedback into consideration in order to identify areas of improvement within the appeals process.

Reporting system

In order to ensure the integrity of our reporting systems, we deploy a combination of automated moderation technology and specifically trained moderators.

When a user reports content on our platform, our automated systems will consider whether it is in violation of our CGs. If a potential violation is found, the automated moderation system will either pass it on to our moderation teams for further review or, if there is a high degree of confidence that the content violates our CGs, remove it automatically. Automated removal is only applied when violations are clear-cut, such as where the content contains nudity or youth safety. We are constantly investing in improving the precision of our automated moderation systems so we can more effectively remove violative content at scale while also reducing the number of incorrect removals.

Where violations are less clear-cut, content will be passed on to our moderation teams for further review, in order to support the fair and consistent review of potentially violative content. As mentioned above, our moderators can take into account additional context and nuance which may not always be picked up by technology. Again, in order to ensure consistency within this process and its overall integrity, we have sought to make our CGs as clear and comprehensive as possible and have put in place robust Quality Assurance processes (including steps such as review of moderation cases, flows, appeals and undertaking Root Cause Analyses).

	We also note that whilst user reports are important, at TikTok we place considerable emphasis on proactive content moderation and so are proud that the vast majority of content is identified and removed proactively before it receives any views or is reported to us.

V. Empowering Users	
Commitment 24 Relevant Signatories commit to inform users whose content or accounts has been subject to enforcement actions (content/accounts labelled, demoted or otherwise enforced on) taken on the basis of violation of policies relevant to this section (as outlined in Measure 18.2), and provide them with the possibility to appeal against the enforcement action at issue and to handle complaints in a timely, diligent, transparent, and objective manner and to reverse the action without undue delay where the complaint is deemed to be founded.	
In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	N/A
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/ No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	As part of our implementation of the DSA, and building on our current tools and processes, we are developing the necessary technical, policy and resource measures required in order to ensure that we are compliant with the content moderation obligations under the DSA - in particular with regard to notifications (Article 17) and appeals (Articles 20 and 21). We intend to implement the relevant additional measures in the Summer of 2023.
Measure 24.1	
QRE 24.1.1	When we take action under our CGs, users are notified by an in-app notification in all EU member states and in an official language of the European Union where their account has been banned or their content has been removed for violating our CGs, as well as where the unverified label has been applied to their video.

Such notifications are provided in near real time after action has been taken (i.e. generally within several seconds or up to a few minutes at most). In relation to content removal, an in-app inbox notification sets out the specific CG policy deemed to have been violated, along with an option for users to submit an appeal by clicking on the "Submit an appeal" button (in bold text on the notification screen) if they believe no violation has occurred. For account bans, a banner notification will inform the user of the account change and give them the option to appeal by clicking on the "Appeal" button within the banner. In relation to the application of the unverified label, a banner notification will inform the user that the label has been applied and will give them the option to "Request a review".

All such appeals raised (i.e. those submitted in relation to content removal, an account ban or the application of the unverified content label) will be queued for review by our specialised human moderators so as to ensure that context is adequately taken into account in reaching a determination. Users can monitor the status and view the results of their appeal within their in-app inbox.

As mentioned above, our users have the ability to share feedback with us to the extent that they don't agree with the result of their appeal. They can do so by using the in-app function which allows them to "report a problem". We are continuously taking user feedback into consideration in order to identify areas of improvement within the appeals process.

In relation to the metrics we have shared below in response to SLI 24.1.1, we note again that misinformation is a nuanced and complex issue to moderate. It's highly contextual, requiring an assessment of accuracy for claims that often evolve over time. Our priority is creating a safe and authentic environment for our users, and we strive to mitigate the harm of misinformation in a timely manner while also providing a fair appeals process.

SLI 24.1.1 - enforcement actions	**Methodology of data measurement:**			
	The number of appeals/overturns is based on the country in which the video being appealed/overturned was posted. These numbers are only related to our harmful misinformation policies. The appeal success rate of videos removed by our harmful misinformation policies is based on the ratio between the number of appeals raised and the number of successful appeals (i.e. overturns).			
	Nr of enforcement actions	Nr of actions appealed	Metrics on results of appeals	Metrics on the duration and effectiveness of the appeal process
Member States				
List actions per member states and languages (see example table above)	**Number of accounts removed banned under our I&A policies**	**Number of appeals of videos removed for**	**Number of successful appeals for violation of harmful**	**Appeal success rate of videos removed for**

		violation of harmful misinformation policy	misinformation policy (i.e. overturns)	violation of harmful misinformation policy
Austria		13	69	8	11.59%
Belgium		8	0	0	0.00%
Bulgaria		8	11	8	72.73%
Croatia		2	1	1	100.00%
Cyprus		0	2	2	100.00%
Czech Republic		60	114	72	63.16%
Denmark		30	9	7	77.78%
Estonia		1	1	0	0.00%
Finland		44	33	15	45.45%
France		586	2,876	1,351	46.97%
Germany		562	2,000	300	15.00%
Greece		23	46	25	54.35%
Hungary		4	81	50	61.73%
Ireland		10	205	147	71.71%
Italy		308	3,792	1,093	28.82%
Latvia		3	1	1	100.00%
Lithuania		1	1	1	100.00%
Luxembourg		0	0	0	0.00%
Malta		0	0	0	0.00%

Netherlands	40	156	37	23.72%
Poland	473	697	324	46.48%
Portugal	37	48	27	56.25%
Romania	15	183	108	59.02%
Slovakia	1	25	19	76.00%
Slovenia	2	3	3	100.00%
Spain	88	1,424	652	45.79%
Sweden	37	302	154	50.99%
Iceland	0	5	4	80.00%
Liechtenstein	0	0	0	0.00%
Norway	51	25	15	60.00%
Total EU	2,356	12,080	4,405	36.47%
Total EEA	2,407	12,110	4,424	36.53%

V. Empowering Users
Commitment 25
In order to help users of private messaging services to identify possible disinformation disseminated through such services, Relevant Signatories that provide messaging applications commit to continue to build and implement features or initiatives that empower users to think critically about information they receive and help them to determine whether it is accurate, without any weakening of encryption and with due regard to the protection of privacy.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No

If yes, list these implementation measures here [short bullet points].	This commitment is not applicable as TikTok is not a messaging app.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No
If yes, which further implementation measures do you plan to put in place in the next 6 months?	This commitment is not applicable as TikTok is not a messaging app.
Measure 25.1	
QRE 25.1.1	**Not committed.** This commitment is not applicable as TikTok is not a messaging app.
SLI 25.1.1	N/A
	N/A
Member States	
List actions per member states and languages (see example table above)	
Measure 25.2	
QRE 25.2.1	**Not committed.** This commitment is not applicable as TikTok is not a messaging app.
SLI 25.2.1 - use of select tools	N/A
	N/A
Member States	
List actions per member states and languages (see example table above)	

VI. Empowering the research community[7]	
Commitment 26	
Relevant Signatories commit to provide access, wherever safe and practicable, to continuous, real-time or near real-time, searchable stable access to non-personal data and anonymised, aggregated, or manifestly-made public data for research purposes on Disinformation through automated means such as APIs or other open and accessible technical solutions allowing the analysis of said data.	
In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	Yes
If yes, list these implementation measures here [short bullet points].	Advanced the build of an API designed to provide researchers with access to relevant data on harmful misinformation.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	- Expanding the build of a dedicated API to provide researchers with access to relevant data on disinformation and engaging with EDMO and any dedicated taskforce for this purpose. - Developing an API for access to our moderation system allowing researchers to evaluate our content moderation systems and examine existing content available on our platform.
Measure 26.1	
QRE 26.1.1	We are in the process of building a dedicated API in order to provide researchers with access to relevant data on disinformation. We are engaging with EDMO on this priority. In parallel, over the past months, we have been working on developing a global and separate transparency API that will provide selected researchers with access to various public and anonymized data from our platform.
QRE 26.1.2	See our response to QRE 26.1.1.

---

[7] Please note that defined terms have the meaning set out in our definitions section here.

SLI 26.1.1 - The uptake of the tools and processes described in Measure 26.1	N/A		
	N/A	N/A	N/A
Member States			
List actions per member states and languages (see example table above)			
Measure 26.2			

QRE 26.2.1	See our response to QRE 26.1.1.
QRE 26.2.2	See our response to QRE 26.1.1.
QRE 26.2.3	See our response to QRE 26.1.1.

SLI 26.2.1 - meaningful metrics on the uptake, swiftness, and acceptance level of the tools and processes in Measure 26.2	N/A					
	N/A	N/A	N/A	N/A	N/A	N/A
Member States						
List actions per member states and languages (see example table above)						
Measure 26.3						

QRE 26.3.1	See our response to QRE 26.1.1.

VI. Empowering the research community

Commitment 27

Relevant Signatories commit to provide vetted researchers with access to data necessary to undertake research on Disinformation by developing, funding, and cooperating with an independent, third-party body that can vet researchers and research proposals.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	Yes
If yes, list these implementation measures here [short bullet points].	● Attended meetings with EDMO to progress measures under this commitment.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	● Actively participating in the taskforce being set up to further this commitment. ● Will continue dedicating resources to engage with EDMO.
Measure 27.1	
QRE 27.1.1	We have been engaging with EDMO as part of this process and are committed to participating in the working group that is being set up in order to put in place the independent third party body that is referred to above, including by nominating a TikTok representative.
Measure 27.2	
QRE 27.2.1	See our response to QRE 27.1.1.
Measure 27.3	
QRE 27.3.1	See our response to QRE 27.1.1.
SLI 27.3.1 - research projects vetted by the independent third-party body	N/A
	N/A
Member States	

List actions per member states and languages (see example table above)	
Measure 27.4	
QRE 27.4.1	We will be launching a pilot phase as part of our efforts to develop an API for researchers as we are keen on engaging with the research community and making sure that our tools, and the data provided through the API, suit their needs. In the meantime, we've asked the members of our Content and Safety Advisory Councils with expertise in various subject matters, including misinformation, to test an early version of the global API for researchers that we are developing in parallel (See QRE 26.1.1). The aim is to gather their feedback on usability and the overall experience of accessing public data through this API.

VI. Empowering the research community
Commitment 28
Relevant Signatories commit to support good faith research into Disinformation that involves their services.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	N/A
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	See the APIs we are in the process of developing set out in our response to Commitment 26.
Measure 28.1	
QRE 28.1.1	TikTok is committed to facilitating research and engaging with the research community.

	We have put in place a **Safety Advisory Council for Europe** that brings together independent leaders from academia and civil society from all around the continent. Each member provides subject matter expertise as they advise on issues relating to users safety, content policies and areas of social relevance. In addition, many of our teams engage with the research community. Our **Outreach & Partnerships Management Team** is dedicated to engaging regularly with, and setting up partnerships with, the academic and research community. This team, together with our subject matter experts within our product team, are central to our fact-checking programme from identifying new partners and onboarding them to regularly meeting with them. More generally, our Trust and Safety teams regularly **consult and engage with the research community**, including on harmful misinformation and deceptive behaviours, when updating or launching new policies or features on our platform. For instance, our teams have engaged with dozens of academics in Europe as part of the development and implementation of our policy related to state affiliated media. During election times, we also consult expert partners to provide us with insights to better understand areas that could potentially amount to election manipulation. We also have many teams committing time to facilitating research. Individuals with backgrounds in product, data science, outreach and legal are working together to build an API to share information on harmful misinformation as well as a global transparency API (see QRE 26.1.1). TikTok representatives also regularly participate in **research-focused events**. Relevant examples from 2022 include the Terrorism and Social Media Conference (Swansea University, June), the Trust and Safety Research Conference (Stanford University, September), and the Association of Internet Researchers Conference (Technological University Dublin, November). At these events we seek to learn from the important work being done by the research community on various topics, which include aspects related to harmful misinformation, as well as share context about our approach, research interests, and opportunities to collaborate.
Measure 28.2	
QRE 28.2.1	We are currently working on developing APIs to allow researchers to access transparent public and anonymized data about content and activity on our platform (see QREs for Commitment 26).
Measure 28.3	
QRE 28.3.1	We are eager to engage and cooperate with EDMO as part of this annual consultation and to receive feedback from the research community on how we facilitate their research projects on our platform.

Measure 28.4	
QRE 28.4.1	We will be happy to provide resources to allow for independent research on harmful misinformation.

VI. Empowering the research community	
Commitment 29 Relevant Signatories commit to conduct research based on transparent methodology and ethical standards, as well as to share datasets, research findings and methodologies with relevant audiences.	
In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	This Commitment relates to Research Organisations.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No
If yes, which further implementation measures do you plan to put in place in the next 6 months?	This Commitment relates to Research Organisations.
Measure 29.1	
QRE 29.1.1	Not committed.
QRE 29.1.2	Not committed.
QRE 29.1.3	Not committed.
SLI 29.1.1 - reach of stakeholders or citizens informed about the outcome of research projects	N/A

	N/A
Member States	
List actions per member states and languages (see example table above)	
Measure 29.2	
QRE 29.2.1	Not committed.
QRE 29.2.2	Not committed.
QRE 29.2.3	Not committed.
SLI 29.2.1	N/A
	N/A
Member States	
List actions per member states and languages (see example table above)	
Measure 29.3	
QRE 29.3.1	Not committed.
SLI 29.3.1 - reach of stakeholders or citizens informed about the outcome of research projects	N/A
	N/A
Member States	
List actions per member states and languages (see example table above)	

VII. Empowering the fact-checking community[8]	
Commitment 30	
Relevant Signatories commit to establish a framework for transparent, structured, open, financially sustainable, and non-discriminatory cooperation between them and the EU fact-checking community regarding resources and support made available to fact-checkers	
In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	Yes
If yes, list these implementation measures here [short bullet points].	- Secured new fact-checking partnerships in Sweden, Hungary, Poland and Romania and expanded our misinformation moderation teams for these countries. - Built a repository of previously fact-checked claims to help our specialised misinformation moderators make swift and accurate decisions. - Conducted our first survey with our partners to review our working relationship and introduced changes. - Increased information sharing with our fact-checking partners about the enforcement action taken as a result of their feedback.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	- Driving forwards our ambitious plans to further expand our fact-checking network, by adding new fact-checking partners and increasing the number of languages our current fact-checking partners provide support in. Over the next year, we will be focusing on setting up fact-checking partnerships in Portugal, Denmark, Greece and Belgium as well as other EU countries. As part of these plans, we will be expanding our misinformation moderation teams. - Continuing to invest in and develop our fact-checking repository. - Continuing to increase the volume of information we share with our fact-checking partners. - Considering where we can make improvements to our partner agreements as part of our annual

---

[8] Please note that defined terms have the meaning set out in our definitions section here.

	contract renewal process.
Measure 30.1	
QRE 30.1.1	Within Europe, we work with 8 fact-checking partners who provide fact-checking coverage in 10 official EU languages. Our partners have teams of fact-checkers who review and verify reported content. Our moderators then use that independent feedback to take action and where appropriate, remove or make ineligible for recommendation false or misleading content or label unverified content. Our agreements with our partners are standardised, meaning the agreements are based on our template master services agreements and consistent of common standards and conditions. The terms of the agreements describe the service the fact-checking partner will provide, namely that their team of fact checkers review, assess and rate video content uploaded to their fact-checking queue. The expected results are also specified e.g., the fact-checkers advise on whether the content may be or contain misinformation and rate it using our classification categories. We can agree that our fact-checker partners provide regular written reports about disinformation trends identified. The agreements also clarify all other key terms including the applicable term and fees and payment arrangements. We also agree with the fact-checking organisations the languages in which they will provide fact-checking services and we can request temporary coverage regarding additional languages if needed. As part of our annual contract renewals process with our partners, we will be reviewing our agreements with our partners and considering whether we can further improve and refine them.
QRE 30.1.2	Within Europe, our IFCN-accredited fact-checking partners are: 1. Agence France Press; 2. Facta.news; 3. Lead Stories; 4. Logically; 5. Newtral; 6. Science Feedback; 7. dpa Deutsche Presse-Agentur; and 8. Teyit. We can, and have, put in place temporary agreements with fact checking partners to provide additional European language coverage for a period in an unfolding crisis. For example, in March 2022, we swiftly put in place temporary fact-checking coverage with Lead Stories for the duration of the Hungarian election (before shifting into permanent coverage in Hungary).

	Globally, we have 15 IFCN-accredited fact-checking partners. We are continuously working to expand our fact-checking network and we keep users updated here.
QRE 30.1.3	We have fact-checking coverage in 10 official European languages (Dutch, English, French, German, Hungarian, Italian, Polish, Romanian, Spanish and Swedish), and, therefore, the spoken language of 15 EEA markets. We can request (and have previously requested) temporary coverage in relation to a number of European languages with a current partner e.g., Hungarian or languages which affect European users, including Azeri, Armenian, Turkish, Russian, Ukrainian and Belarusian. We can also put a temporary arrangement in place with a new partner. In terms of global fact-checking initiatives, we currently cover 46 languages and assess content in 71 markets, thereby improving the overall integrity of the service and benefiting European users. As mentioned, we have ambitious plans to continue the expansion of our fact-checking network in Europe and globally. In order to effectively scale the feedback provided by our fact-checkers globally, we have implemented the measures listed below. ● **Fact-checking repository.** In the last year we have built a database of claims which have previously been fact-checked by our EU and global fact checking partners. ● **Trends reports.** Our fact-checking partners can provide us with regular reports identifying general misinformation trends observed on our platform and across the industry generally, including new/changing industry or market trends, events or topics that generated particular misinformation or disinformation. ● **Proactive detection by our fact-checking partners.** Certain of our fact-checking partners are authorised to proactively identify content that may constitute harmful misinformation on our platform and suggest prominent misinformation that is circulating on other social media platforms or websites that may benefit from verification. ● **Fact-checking guidelines.** We create guidelines and trending topic reminders for our moderators on the basis of previous fact-checking assessments. This ensures our moderation teams leverage the insights from our fact-checking partners and helps our moderators make swift and accurate decisions on flagged content regardless of the language in which the original claim was made. Our specialised misinformation moderation teams have direct access to these tools and measures, which enables them to more swiftly, easily and accurately take action on violating content across Europe and globally. We are also in the process of building and testing models which may allow for the output of these measures to be used to update the machine learning models we use in proactive detection, learning, over time, to search for similar content which can be proactively recalled into our moderation system for review.

	Leveraging the fact-checking output in this way would enable us to further increase the positive impact of our fact checking programme.
SLI 30.1.1 - Member States and languages covered by agreements with the fact-checking organisations	See our response to QRE 30.1.3.
Member States	
List actions per member states and languages (see example table above)	
Measure 30.2	
QRE 30.2.1	Our agreements with our fact-checking partners are standardised, meaning the agreements are based on our template master services agreements and consistent of common standards and conditions. These agreements, as with all of our agreements, must meet the ethical and professional standards we set internally including containing anti-bribery and corruption provisions. Our partners are compensated in a fair, transparent way based on the work done by them. We are clear with our partners on the compensation they should expect from us from the outset in the relevant agreement. Our fact-checking partners then invoice us on a monthly basis based on work done. All of our fact-checking partners are independent organisations, which are certified through the non-partisan IFCN. Our agreements with them explicitly state that the fact-checkers are non-exclusive, independent contractors of TikTok who retain editorial independence in relation to the fact-checking, and that the services shall be performed in a professional manner and in accordance with the highest standards in the industry. Our processes are also set up to ensure our fact-checking partners independence. Our partners access flagged content through an exclusive dashboard for their use and provide their assessment of the accuracy of the content by providing a rating. Fact-checkers will do so independently from us, and their review may include calling sources, consulting public data or authenticating videos and images. To facilitate transparency and openness with our fact-checking partners, we regularly meet them and provide data regarding their feedback and also conduct surveys with them.
QRE 30.2.2	We meet regularly with our fact-checking partners and have an ongoing dialogue with them about how our partnership is working and evolving. We recently surveyed our fact-checking partners to encourage feedback about what we are doing well and how we could improve, and as a result of that, we introduced some changes.

	We have recently shared with our fact-checking partners detailed data on the enforcement action that was taken regarding the videos they have reviewed, and are collaborating with them to understand what additional information may be useful for us to share going forward.
QRE 30.2.3	This provision is not relevant to TikTok, only to fact-checking organisations.
Measure 30.3	
QRE 30.3.1	Given our fact-checking partners are all IFCN-accredited, our fact-checking partners already engage in some informal cross-border collaboration through that network. In addition, we are working with our partners to understand how we may be able to facilitate collaboration.
Measure 30.4	
QRE 30.4.1	We are in regular dialogue with EDMO and the EFCSN on these and other issues. We continue to be open to discussing and exploring what further progress can be made on these points.

VII. Empowering the fact-checking community
Commitment 31
Relevant Signatories commit to integrate, showcase, or otherwise consistently use fact-checkers' work in their platforms' services, processes, and contents; with full coverage of all Member States and languages.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	Yes
If yes, list these implementation measures here [short bullet points].	- New contracts with fact-checking partners in Sweden, Hungary, Poland and Romania. - Developing our fact-checking repository to ensure our teams and systems leverage the full scope of insights our fact-checking partners submitted to TikTok (regardless of the original language of the relevant content).

Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	- Expanding fact-checking coverage in additional EU member states and their languages. Over the next year, we are focusing on setting up fact-checking partnerships in Portugal, Denmark, Greece and Belgium as well as other EU countries. - Continuing to invest in our fact-checking repository and tools for fact-checking repository.
Measure 31.1	
Measure 31.2	
QRE 31.1.1	We see harmful misinformation as different from other content issues. Context and fact-checking are critical to consistently and accurately enforcing our harmful misinformation policies, which is why we work with 8 fact-checking partners in Europe, covering 10 official European languages.

While we use machine learning models to help detect potential misinformation, our approach is to have our specialised misinformation moderators assess, confirm, and take action on harmful misinformation. This includes direct access to our fact-checking partners who help assess the accuracy of content. Our fact-checking partners are involved in our moderation process in three ways:

(i) a moderator sends a video to fact-checkers for review and their assessment of the accuracy of the content by providing a rating. Fact-checkers will do so independently from us, and their review may include calling sources, consulting public data, authenticating videos and images, and more.

Out of an abundance of caution, while content is being fact-checked or when content can't be substantiated through fact-checking, we may significantly reduce the content's distribution so that fewer people see it. Fact-checkers ultimately do not take action on the content directly. The moderator will instead take into account the fact-checkers' feedback on the accuracy of the content when deciding whether the content violates our CGs and what action to take.

(ii) contributing to our global database of previously fact-checked claims to help our misinformation moderators make swift and accurate decisions.

(iii) a proactive detection programme with our fact-checkers who flag new and evolving claims they're seeing on our platform. This enables our moderators to quickly assess these claims and remove violations.

In addition, we use fact-checking feedback to provide additional context to users about certain content. As mentioned, when sometimes our fact checking partners conclude that the fact-check is inconclusive or content is not able to be confirmed, (which is especially common during unfolding events or crises), we |

	inform viewers via a banner when we identify a video with unverified content in an effort to raise users' awareness about the credibility of the content and to reduce sharing. The video may also become ineligible for recommendation into anyone's For You feed to limit the spread of potentially misleading information.
	In relation to the metrics we have shared below, we note again that we are continuing to expand our fact-checking program both in terms of the EEA countries in which we have fact-checking partners (and thus onboarding new partners) and the scaling up the volume of claims we fact check and content our current partners are reviewing. As aforementioned, over the next year, we are focusing on setting up fact-checking partnerships in Portugal, Denmark, Greece and Belgium as well as other EU countries.
	As we build momentum with our fact checking partners, the metrics will also reflect this progress.

SLI 31.1.1 - use of fact-checks	**Methodology of data measurement:**		
	The number of fact checked videos is based on the number of videos that have been sent for review to one of our fact-checking partners in the relevant territory.		
	Nr of fact-checked articles published	N/A	N/A	N/A
Member States				
List actions per member states and languages (see example table above)	**Number of fact checked videos**			
Austria	17			
Belgium	0			
Bulgaria	0			
Croatia	0			
Cyprus	0			
Czech Republic	0			
Denmark	0			

Estonia	0			
Finland	0			
France	962			
Germany	157			
Greece	0			
Hungary	5			
Ireland	28			
Italy	347			
Latvia	0			
Lithuania	0			
Luxembourg	0			
Malta	0			
Netherlands	50			
Poland	44			
Portugal	0			
Romania	478			
Slovakia	0			
Slovenia	0			
Spain	270			
Sweden	448			

Iceland	0			
Liechtenstein	0			
Norway	0			
Total EU	2,806			
Total EEA	2,806			

SLI 31.1.2 - impact of actions taken	**Methodology of data measurement:**			
	The number of videos removed because of policy guidelines, known misinformation trends and our knowledge based repository is based on the country in which the video was posted. These metrics correspond to the numbers of removals under the harmful misinformation policy since all of its enforcement are based on the policy guidelines, known misinformation trends and knowledge based repository.			
	N/A		Impact of said measures on user interactions with information labelled as false or misleading	Other
Member States				
List actions per member states and languages (see example table above)	N/A	N/A	Number of videos removed as a result of a fact checking assessment	Number of videos removed because of policy guidelines, known misinformation trends and knowledge based repository
Austria			0	174
Belgium			0	4
Bulgaria			0	38

Croatia			0		7
Cyprus			0		7
Czech Republic			0		220
Denmark			0		29
Estonia			0		1
Finland			0		73
France			250		9,442
Germany			45		5,855
Greece			0		207
Hungary			3		202
Ireland			0		485
Italy			94		10,491
Latvia			0		1
Lithuania			0		3
Luxembourg			0		0
Malta			0		0
Netherlands			0		401
Poland			19		2,006

Portugal			0		123	
Romania			145		912	
Slovakia			0		36	
Slovenia			0		4	
Spain			67		4,918	
Sweden			81		868	
Iceland			0		11	
Liechtenstein			0		0	
Norway			0		59	
Total EU			704		36,507	
Total EEA			704		36,577	

SLI 31.1.3 – Quantitative information used for contextualisation for the SLIs 31.1.1 / 31.1.2.	**Methodology of data measurement:** The metric we have provided demonstrates the % of videos which have been removed as a result of the fact checking assessment, in comparison to the total number of videos removed because of violation of our harmful misinformation policy.
Member States	
List actions per member states and languages (see example table above)	**Number of videos removed as a result of a fact checking assessment / number of removals under harmful misinformation policy**
Total EU	1.92%

Measure 31.3	
QRE 31.3.1	We are in regular engagement with EDMO on this priority. We have committed to participate in the taskforce made up of the relevant signatories' representatives that EDMO has created for this purpose.
Measure 31.4	
QRE 31.4.1	As above.

VII. Empowering the fact-checking community
Commitment 32 Relevant Signatories commit to provide fact-checkers with prompt, and whenever possible automated, access to information that is pertinent to help them to maximise the quality and impact of fact-checking, as defined in a framework to be designed in coordination with EDMO and an elected body representative of the independent European fact-checking organisations.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	Yes
If yes, list these implementation measures here [short bullet points].	- We launched a new landing page dedicated to highlighting our fact-checking partners and how we work with them (see here). - Our Global Head of Trust and Safety published a blog post sharing more about our integrity systems and specifically how we partner with fact-checkers to address misinformation. - We started giving partners deeper insights into their impact as part of the fact-checking program.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	- Sharing more regular updates with our fact-checking partners regarding their impact in the program. - Investing more resources in the tools used to fact-check on our platforms.
Measure 32.1	

Measure 32.2	
QRE 32.1.1	Our fact-checking partners access content which has been flagged for review through a dashboard made available for their exclusive use. The dashboard shows our fact-checkers certain quantitative information about the services they provide, including the number of videos queued for assessment at any one time, as well as the time the review has taken. Fact-checkers can also use the dashboard to see the rating they applied to videos they have previously assessed. We have also recently shared more data with our partners at our regular meetings to help them quantify the impact of the fact-checked content over time e.g., the number of videos removed as a result of their feedback. We continue to work with our partners to understand further what data it would be helpful for us to share with them and in what form.
SLI 32.1.1 - use of the interfaces and other tools	**Methodology of data measurement:** N/A. As mentioned in our response to QRE 32.1.1, the dashboard we currently share with our partners only contains high level quantitative information about the services they provide, including the number of videos queued for assessment at any one time, as well as the time the review has taken. We are working with our fact checking partners to understand what further data it would be helpful for us to share with them.

	N/A	N/A	N/A
Member States			
List actions per member states and languages (see example table above)			

Measure 32.3	
QRE 32.3.1	We are committed to participate in the taskforce made up of the relevant signatories' representatives that is being set up for this purpose. Meanwhile we are also engaging with EDMO pro-actively on this commitment.

VII. Empowering the fact-checking community

Commitment 33

Relevant Signatories (i.e. fact-checking organisations) commit to operate on the basis of strict ethical and transparency rules, and to protect their independence.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No	
If yes, list these implementation measures here [short bullet points].	This Commitment and Measure relates to fact checking organisations.	
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No	
If yes, which further implementation measures do you plan to put in place in the next 6 months?	This Commitment and Measure relates to fact checking organisations.	
Measure 33.1		
QRE 33.1.1	**Not committed**. This Commitment and Measure relates to Fact-Checking organisations.	
SLI 33.1.1 - number of European fact-checkers that are IFCN-certified	N/A	
	N/A	N/A
Member States		
List actions per member states and languages (see example table above)		

VIII. Transparency Centre[9]

Commitment 34

---

[9] Please note that defined terms have the meaning set out in our definitions section here.

To ensure transparency and accountability around the implementation of this Code, Relevant Signatories commit to set up and maintain a publicly available common Transparency Centre website	
In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	Yes
If yes, list these implementation measures here [short bullet points].	We have been an active participant in the subgroup that has successfully launched the common Transparency Centre this year.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	We are committed to maintaining the common Transparency Centre and to continuing our active participation in the Transparency Subgroup.
Measure 34.1	
Measure 34.2	
Measure 34.3	
Measure 34.4	
Measure 34.5	

VIII. Transparency Centre
Commitment 35
Signatories commit to ensure that the Transparency Centre contains all the relevant information related to the implementation of the Code's Commitments and Measures and that this information is presented in an easy-to-understand manner, per service, and is easily searchable.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	Yes
If yes, list these implementation measures here [short bullet points].	Through our participation in the subgroup that has successfully launched the common Transparency Centre this year, we have ensured that the Centre will allow the general public to access general information about the Code as well as the underlying reports (and for the Centre to be navigated both by commitment and signatory).
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	We are committed to maintaining the common Transparency Centre and to continuing our active participation in the Transparency Subgroup.
Measure 35.1	
Measure 35.2	
Measure 35.3	
Measure 35.4	
Measure 35.5	
Measure 35.6	

VIII. Transparency Centre
Commitment 36

Signatories commit to updating the relevant information contained in the Transparency Centre in a timely and complete manner. |

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	Not applicable for the baseline report.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	We are committed to maintaining the common Transparency Centre and, as part of this, updating the requisite information in a timely and complete manner.
Measure 36.1	
Measure 36.2	
Measure 36.3	
QRE 36.1.1 (**for the Commitments 34-36**)	We are pleased to confirm that we have been an active participant in the subgroup that has successfully launched the common Transparency Centre this year. We have assisted with establishing the website's requirements, selecting a vendor to build the website and overseeing the development of the website's key functionalities and interface. We have ensured that the Centre will allow the general public to access general information about the Code as well as the underlying reports (and for the Centre to be navigated both by commitment and signatory). Each signatory will be responsible for ensuring that the information they upload to the website is correct and accurate. Entities interested in joining the Code's Task-force will be able to sign up through a dedicated online application form on the website.
QRE 36.1.2 (**for the Commitments 34-36**)	Not applicable for the baseline report.
SLI 36.1.1 - (for Measures 34 and 36) meaningful quantitative information on the usage of the Transparency Centre, such as the average monthly visits of the webpage.	N/A
	N/A
Member States	
List actions per member states and languages (see example table above)	

IX. Permanent Task-Force[10]

Commitment 37

Signatories commit to participate in the permanent Task-force. The Task-force includes the Signatories of the Code and representatives from EDMO and ERGA. It is chaired by the European Commission, and includes representatives of the European External Action Service (EEAS). The Task-force can also invite relevant experts as observers to support its work. Decisions of the Task-force are made by consensus.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	N/A
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No
If yes, which further implementation measures do you plan to put in place in the next 6 months?	N/A - no current plans.
Measure 37.1	
Measure 37.2	
Measure 37.3	
Measure 37.4	
Measure 37.5	

---

[10] Please note that defined terms have the meaning set out in our definitions section here.

Measure 37.6	
QRE 37.6.1	We will continue to participate in the Permanent Task-force and will assess needs in the context of ongoing work undertaken by the group.

X. Monitoring of Code[11]	
Commitment 38 The Signatories commit to dedicate adequate financial and human resources and put in place appropriate internal processes to ensure the implementation of their commitments under the Code.	
In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	Yes
If yes, list these implementation measures here [short bullet points].	New implementation measures to address the obligations within the Code.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Planned implementation measures in view of our ongoing obligations under the Code.
Measure 38.1	
QRE 38.1.1.	TikTok has assigned the highest priority level to the Code, which means that we have, and will continue to have, appropriate resources in place to meet our commitments and compliance.

---

[11] Please note that defined terms have the meaning set out in our definitions section here.

	Given the breadth of the Code and the commitments therein, our work spans multiple teams, including Trust and Safety, Legal, Monetisation Integrity, Product and Public Policy. Teams across the globe are deployed to ensure that we meet our commitments and compliance with the notable involvement of our Trust and Safety Leadership team which is based in Dublin, Ireland. TikTok has thousands of content moderators globally and across the European region, as we recognise the importance of local knowledge and expertise as we work to ensure online safety for our users. We take a similar approach to our third party partnerships.

X. Monitoring of Code

Commitment 39

Signatories commit to provide to the European Commission, within 1 month after the end of the implementation period (6 months after this Code's signature) the baseline reports as set out in the Preamble.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	Yes
If yes, list these implementation measures here [short bullet points].	We have shared this baseline report with the Commission in accordance with the relevant timeframes.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No
If yes, which further implementation measures do you plan to put in place in the next 6 months?	N/A

X. Monitoring of Code

Commitment 40	
Signatories commit to provide regular reporting on Service Level Indicators (SLIs) and Qualitative Reporting Elements (QREs). The reports and data provided should allow for a thorough assessment of the extent of the implementation of the Code's Commitments and Measures by each Signatory, service and at Member State level.	
In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	Yes
If yes, list these implementation measures here [short bullet points].	We reported on the SLIs and QREs relevant to the Commitments we signed-up to within this baseline report to the extent possible at this time.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	We are committed to reporting on the SLIs and QREs relevant to the Commitments we signed-up to in accordance with the timeframes agreed between the taskforce.
Measure 40.1	
Measure 40.2	
Measure 40.3	
Measure 40.4	
Measure 40.5	
Measure 40.6	

X. Monitoring of Code

Commitment 41

Signatories commit to work within the Task-force towards developing Structural Indicators, and publish a first set of them within 9 months from the signature of this Code; and to publish an initial measurement alongside their first full report. To achieve this goal, Signatories commit to support their implementation, including the testing and adapting of the initial set of Structural Indicators agreed in this Code. This, in order to assess the effectiveness of the Code in reducing the spread of online disinformation for each of the relevant Signatories, and for the entire online ecosystem in the EU and at Member State level. Signatories will collaborate with relevant actors in that regard, including ERGA and EDMO.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	Yes
If yes, list these implementation measures here [short bullet points].	We have been an active participant in the sub-group dedicated to developing Structural Indicators.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	We are committed to working with the sub-group to develop and implement the Structural Indicators in accordance with the agreed timeframes
Measure 41.1	
Measure 41.2	
Measure 41.3	

X. Monitoring of Code

Commitment 42

Relevant Signatories commit to provide, in special situations like elections or crisis, upon request of the European Commission, proportionate and appropriate information and data, including ad-hoc specific reports and specific chapters within the regular monitoring, in accordance with the rapid response system established by the Taskforce.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	Yes
If yes, list these implementation measures here [short bullet points].	We have been an active participant in the Crisis Response Sub-group and have shared reports specific to the War of aggression by Russia on Ukraine and the Covid-19 pandemic along with this baseline report.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	No
If yes, which further implementation measures do you plan to put in place in the next 6 months?	Not applicable.

X. Monitoring of Code

Commitment 43

Signatories commit to produce reports and provide data following the harmonised reporting templates and refined methodology for reporting and data disclosure, as agreed in the Task-force.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	Yes
If yes, list these implementation measures here [short bullet points].	New implementation measures to address the various commitments throughout the Code, as reported on in the relevant sections of the report templates.
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes

If yes, which further implementation measures do you plan to put in place in the next 6 months?	We have identified and called out areas for ongoing progress/work in various places within the reports.

X. Monitoring of Code

Commitment 44
Relevant Signatories that are providers of Very Large Online Platforms commit, seeking alignment with the DSA, to be audited at their own expense, for their compliance with the commitments undertaken pursuant to this Code. Audits should be performed by organisations, independent from, and without conflict of interest with, the provider of the Very Large Online Platform concerned. Such organisations shall have proven expertise in the area of disinformation, appropriate technical competence and capabilities and have proven objectivity and professional ethics, based in particular on adherence to auditing standards and guidelines.

In line with this commitment, did you deploy new implementation measures (e.g. changes to your terms of service, new tools, new policies, etc)? [Yes/No]	No
If yes, list these implementation measures here [short bullet points].	N/A
Do you plan to put further implementation measures in place in the next 6 months to substantially improve the maturity of the implementation of this commitment? [Yes/No]	Yes
If yes, which further implementation measures do you plan to put in place in the next 6 months?	TikTok is committed to meeting its obligations under the DSA, including the independent audit obligations. TikTok is taking steps to ensure readiness for its obligations in this regard and is also awaiting the relevant guidance from the Commission in order to finalise its compliance approach.

Reporting on the service's response during a period of crisis

War of aggression by Russia on Ukraine

Overview of the main threats observed, such as crisis related disinformation campaigns, spread of misinformation, coordinated manipulative behaviours, malicious use of advertising products, involvement of foreign state actors, etc.

The war of aggression by Russia on Ukraine is devastating, and, as a platform, has challenged us to confront a complex and rapidly changing environment. At TikTok, the safety of our people and community is of paramount importance and we work continuously to safeguard our platform. We have set out below some of the main threats we have observed on our platform in relation to the spread of harmful misinformation related to the war and note that we are committed to taking action to prevent such content from being shared in this context.

At the beginning of the war, as military actions against Ukraine commenced, we witnessed false or unconfirmed claims about specific attacks and events, the development or use of weapons, the involvement of specific countries in the conflict and denial of the existence of a country. In addition, there were also cases of content that was repurposed in a misleading way including from video games or unrelated footage from past events. As detailed below, we take a multi-faceted approach to understanding and removing misleading stories and, when it comes to addressing misinformation, we apply the Integrity & Authenticity policies (**I&A policies**) in our Community Guidelines (**CGs**) and we will take action on such content from our platform. We support our moderation teams with detailed policy guidance and direction when moderating on crisis related misinformation using our misinformation policies, including providing case banks of harmful misinformation claims to support their moderation work.

Our I&A policies do not allow coordinated attempts to influence or sway public opinion while also misleading individuals, our community, or our systems about an account's identity, approximate location, relationships, popularity, or purpose. We have specifically-trained teams which are on high alert to investigate and detect covert influence operations (**CIO**) on our platform. We are able to confirm that between July and September 2022, we identified and removed two distinct networks targeting discourse related to the war and 1,704 associated accounts for coordinated efforts to influence public opinion and mislead users about their identities. We published this information voluntarily within our most recently published transparency report, here.

We know that CIO will continue to evolve in response to our detection and networks may attempt to reestablish a presence on our platform, which is why we continually seek to strengthen our policies and enforcement actions in order to protect our community against new types of harmful misinformation and inauthentic behaviours. We have strengthened our efforts to detect suspected accounts impersonating high profile individuals/entities in the affected regions in order to avoid the spread of misinformation or any kind of misleading claims.

Lastly, whilst we do not allow political advertising on our platform and have taken a strict position not to enable associated monetisation, we note that our monetisation integrity content teams observed a sharp rise in attempts to post ads related to political/war content during the initial period of the conflict. Our risk control methods ensured there was no corresponding increase in violating activity going live on the platform. Since the initial period of the conflict, these numbers have significantly dropped.

Executive summary of the company's main strategies and actions taken to mitigate the identified threats and react to the crisis:

We are continually working hard to ensure that TikTok is a source of reliable and safe information and recognise the heightened risk and impact of misleading information during a time of crisis.

We employ a layered approach to detecting harmful misinformation which is in violation of our CGs. In order to ensure coverage which is comprehensive and responsive to the war, we have invested in Russian and Ukrainian native language speakers for content moderation, as we recognise the importance of local cultural and linguistic context within the process.

Working closely with our fact-checking partners has been a crucial part of our approach to removing harmful misinformation from our platform. Our fact-checking programme includes coverage of Russian, Ukrainian and Belarusian, and, since the beginning of the war, we have leveraged Ukrainian-speaking reporters who are connected with some of our existing fact checking partners to further support our fact-checking efforts in Ukraine specifically. Collaborating with certain of our fact-checking partners to receive advance warning of emerging misinformation narratives has facilitated proactive responses against high-harm trends and has ensured that our moderation teams have up-to-date guidance.

Disrupting CIO networks has also been high priority work for us in the context of the crisis. We are able to confirm that between July and September 2022, we quickly identified and removed 2 distinct networks targeting discourse about the war, with 1,704 associated accounts globally, for coordinated efforts to influence public opinion and mislead users about their identities, in breach of our CG policies regarding deceptive behaviours. We published this information voluntarily within our most recently published transparency report, here.

Since the early stages of the war, we have restricted access to content from a number of Russian state affiliated media entities in the EU, Iceland and Liechtenstein. We have also expedited the rollout of our state affiliated media policy to bring viewers context to evaluate the content they consume on our platform and labels have since applied to content posted by the accounts of such entities in Russia, Ukraine and Belarus. Since January 2023, we have expanded the use of the label to media entities across 40 countries (publicly disclosed here).

Whilst we have long prohibited political advertising on our platform, we have taken measures to further mitigate the risk of monetization off the back of the war by disallowing Ukrainian and Russian advertisers from targeting EU markets and removing Livestream videos originating in Russia and Ukraine from the For You feed of users located in the EU.

Lastly, recognising the importance of proactive measures which are aimed at improving our users' digital literacy and increasing the prominence of authoritative information, we have created and promoted an in-app digital literacy campaign specific to information about the war.

Best practices identified for future crisis situations:

- Working closely with our fact-checking partners to receive advance warning of emerging misinformation narratives. This has facilitated proactive responses against high-harm trends and has ensured that our moderation teams have up-to-date guidance.
- Dedicating resources to increasing the number of in-app measures that show users additional context on certain content or redirect them to authoritative information.
- Rolling out media literacy campaigns in partnership with trusted organisations, focusing on addressing disinformation.

Future measures planned within the next 6 months:

- Driving forwards our ambitious plans to further expand our fact-checking network, by adding new fact-checking partners and increasing the number of languages our current fact-checking partners provide support in. Over the next year, we will be focusing on setting up fact-checking partnerships in Portugal, Denmark, Greece and Belgium as well as other EU countries. As part of these plans, we will be expanding our misinformation moderation teams.

- Continuing to work with our fact-checking partners to identify specific disinformation trends in countries and develop tailored, localised media literacy campaigns to tackle those trends and developing our fact-checking repository to ensure our teams and systems leverage the full scope of insights our fact-checking partners submitted to TikTok.
- Further expanding the application of state-controlled media labels to additional countries, noting that since January 2023, we have rolled-out the use of the label to media entities across 40 countries (publicly disclosed here).
- In partnership with trusted organisations, rolling-out a media literacy campaign to address disinformation related to the war on Ukraine in certain Eastern European countries, which will aim to improve the digital literacy of our users.

Changes in Policy Framework		
	Policies	**Rationale**
Policies newly introduced for addressing the crisis	Since the beginning of the reporting period (i.e. October 2022 - December 2022), we have been able to rely on our existing, robust I&A policies and as such have not needed to introduce new policies for the purposes of addressing the crisis.	In the context of the war, we have been able to rely on our I&A policies, which are our first line of defence in combating harmful misinformation and deceptive behaviours on our platform. These guidelines make clear to our users what content we remove or make ineligible for the For You feed when they pose a risk of harm to our users or the wider public. We have also supported our moderation teams with detailed policy guidance and direction when moderating on war-related harmful misinformation using existing policies. We have specialist teams within our Trust and Safety department dedicated to the policy issue of Integrity and Authenticity, including within the areas of product and policy. Our experienced subject matter experts on Integrity and Authenticity continually keep these policies under review and collaborate with external partners and experts when understanding whether updates are required. When situations such as the war in Ukraine arise, these teams work to ensure that appropriate guidance is developed so that the I&A policies are applied in an effective manner in respect of content relating to the relevant crisis (in this case, the war). This includes issuing detailed policy guidance and direction, including providing case banks on harmful misinformation claims to support moderation teams.
Policies adapted for addressing the crisis	Since the beginning of the reporting period, we have been able to rely on the I&A policies, which are an effective basis for tackling content related to the war. As such, we have not adapted our Terms of Service or CGs for addressing the crisis.	See rationale above.
Actions to mitigate the crisis impact on the service		

Type of mitigation	Intervention or action (short summary)	Intervention or action (explanation and implementation)	Impact metrics *NB: the impact metrics we are sharing within this report are based on proxies and, as such, should be regarded as an approximation. Our reporting period for the purpose of these metrics is October - December 2022.*
Actions taken against dis- and misinformation content (for example deamplification, labelling, removal etc.)	**<u>Removing harmful misinformation from our platform</u>** Since the outbreak of the war, we have removed tens of thousands of videos containing harmful misinformation related to the war from our platform. We place significant emphasis on proactive content moderation at TikTok, and are proud that we remove the vast majority of violative videos before they are reported to us by users or other third parties.	We take action to remove accounts or content which contain false or misleading claims that may cause harm to individuals, our community, or the larger public, regardless of intent. In conflict environments, such information may include content that is repurposed from past conflicts, content which makes false claims about specific events, or incites panic. In certain circumstances, we may also reduce the prominence of such content. We employ a layered approach to misinformation detection, leveraging multiple overlapping strategies to ensure comprehensive and responsive coverage. We place considerable emphasis on proactive content moderation strategies in order to remove harmful misinformation which is in violation of our policies before it is reported to us by users or third parties. In the context of the crisis, we are proud to have proactively removed tens of thousands of videos containing harmful misinformation related to the war. We have been able to do this through a combination of automated review, - human level content moderation, carrying out targeted sweeps of certain types of content (e.g. hashtags/sensitive keyword lists) as well as working closely with our fact-checking partners and responding to emerging trends they identify. We have also invested heavily in Russian and Ukrainian native language speakers for content moderation, as we recognise the importance of local cultural and linguistic context in the misinformation moderation process.	**Number of videos removed because of violation of misinformation policy with a proxy related to the War of Aggression by Russia on Ukraine - 1292** **Number of proactive removals of videos removed because of violation of misinformation policy with a proxy related to the War of Aggression by Russia on Ukraine - 1027**
	<u>Applying our state controlled media label</u>	In the EU, Iceland and Liechtenstein, we have taken steps to restrict access to content from the following entities: Russia Today; Sputnik; Rossiya RTR/RTR Planeta; Rossiya 24/Russia 24; and TV Centre	**Number of videos tagged with the state**

	We have restricted access to certain state-affiliated media entities and expedited the roll-out of our state-affiliated media policy and label in order to bring viewers context to evaluate content shared by such Russian, Ukrainian and Belarusian entities.	International. In the wake of the crisis, we expedited the roll out of our state-affiliated media policy in order to bring viewers context to evaluate the content they consume on our platform in relation to the war. In accordance with the policy, we piloted the application of prominent state-controlled media labels to content from such Russian, Ukrainian and Belarusian media entities. Since January 2023, we have expanded the use of the label to media entities across 40 countries (publicly disclosed here). Where the label is applied to content, users across the EEA are automatically shown a full screen pop-up containing information about what the label means and inviting the user to click on "learn more" and be redirected to an in-app page, which explains why the content has been labelled as state-controlled media.	**controlled media label for Russia, Belarus and Ukraine - 3974** **Number of impressions of the state controlled media label for Russia, Belarus and Ukraine - 202004941**
Promotion of authoritative information, including via recommender systems and products and features such as banners and panels	**Not proactively promoting news-type content to our users** We do not actively promote news-type content, such as news related to the war, on our platform. Our recommender system is personalised based on user engagement, so if a user engages with news-type content, it may be the case that more of such content will appear over time, but it will not be actively promoted by us.	TikTok is primarily an entertainment platform. It may be the case that a user's For You feed shows more news-type content (such as news content relating to the war) over time if that is content that a user is actively looking for and engaging with, but we do not actively promote news content to users on TikTok. Further, TikTok's recommender system delivers a personalised feed of content based on a user's engagement behaviour with video content types. Every user's feed will look different, according to their own preferences, which are demonstrated through 'likes', watch duration and other related metrics. For more information on how our recommendation system works, see here.	N/A
	Applying our unverified content label to make content ineligible for recommendation Where our specialised misinformation moderators or fact-checking partners determine that content is not able to be verified at the given time (which is common during an unfolding crisis), we apply our unverified content label to the content to encourage users to consider the	Verifying certain information during dynamic and fast moving events such as a war can be challenging and our moderators and fact-checkers cannot always conclusively determine whether content is indeed harmful misinformation, in violation of our CGs. Therefore, in order to minimise risk, where our fact-checkers or our specialised moderators do not have enough information to verify content which may potentially be misleading, we apply our unverified content label to inform users the content has been reviewed but cannot be conclusively validated. The goal is to raise users' awareness about the credibility of the content and to reduce sharing (see screenshots here). Our unverified content label is available to users in 21 EU official languages (plus, for EEA users, Norwegian and, as the spoken	N/A

	reliability or source of the content. The application of the label will also result in the content becoming ineligible for recommendation in order to limit the spread of misleading information.	language of Liechtenstein, German). Where the banner is applied, the content will also become <u>ineligible for recommendation</u> into anyone's For You feed to limit the spread of information relating to unfolding events where details are still developing and which may potentially be misleading. We are working to improve the accuracy of the data metrics in respect of not recommend moderation actions and will include these in future reports once that work is complete.	
	<u>Removing Russian and Ukrainian Livestream content from the For You feed of EU users</u> We have taken action to remove Livestream videos originating in Russia and Ukraine from the For You feed of users located in the EU and suspend new content creation from our video service in Russia in light of their new "fake news" law.	We have removed Livestream videos originating in Russia and Ukraine from the For You feed of users located in the EU. This has been done in order to minimise the risk of harmful content appearing in a user's feed and given the uptick of Livestreams from that region since the outbreak of the war. We also swiftly took action to suspend new content creation from our video service in Russia in order to review the safety implications of the "fake news" law which was passed by the Russian parliament in March 2022.	**N/A**
	<u>Creating localised media literacy campaigns</u> We launched localised media literacy campaigns relating to the crisis in order to raise awareness amongst our users. We promoted the campaign through a combination of our in-app intervention tools in order to ensure that authoritative information is promoted to our users.	We recognise the importance of proactive measures which are aimed at improving our users' digital literacy and increasing the prominence of authoritative information. As such, we promoted a localised in-app War in Ukraine <u>digital literacy campaign</u> specific to information about the crisis, which encouraged users to verify and evaluate information and provided links to United Nations webpages to facilitate user donations. We used a combination of our in-app intervention tools to bolster and promote the campaign (including video tags, search interventions and redirecting users to trusted resources around the crisis). The use of our intervention tools in relation to campaigns such as our <u>War in Ukraine campaign</u> ensures that authoritative information is promoted to our users. In the coming months, we will be rolling out a renewed campaign in certain Eastern European countries which aims to address	*NB: We implemented localised versions of the digital literacy campaign in 18 European countries, however, do not have engagement data available. We are working on resolving this for future campaigns.*

		disinformation relating to the war and improve the digital literacy of users.	
Cooperation with independent fact-checkers in the crisis context, including coverage in the EU	**Ensuring fact-checking coverage** Our fact checking efforts cover Russian, Ukrainian, Belarusian and all major European languages (including 10 official European languages as well as a number of other languages which affect European users).	We see harmful misinformation as different from other content issues. Context and fact-checking are critical to consistently and accurately enforcing our harmful misinformation policies, which is why we have ensured that, in the context of the crisis, our fact-checking programme covers Russian, Ukrainian and Belarusian. More generally, we work with 8 fact-checking partners in Europe, covering the spoken language of fifteen EEA markets (as well as a number of other languages which affect European users). We have ambitious plans to further expand our network. To further support our fact-checking efforts in Ukraine specifically, we have been leveraging additional Ukrainian-speaking reporters who are connected with some of our existing fact checking partners. Note that we are continuing to expand our fact-checking program both in terms of the EEA countries in which we have fact-checking partners (and thus onboarding new partners) and the scaling up the volume of claims we fact-check and content our current partners are reviewing. As mentioned above, over the next year, we are focusing on setting up fact-checking partnerships in Portugal, Denmark, Greece and Belgium as well as other EU countries. As we build momentum with our fact checking partners, our metrics will reflect this progress.	**Number of fact checked videos with a proxy related to the War of Aggression by Russia on Ukraine - 90** **Number of videos removed as a result of a fact checking assessment with words related to RU/UA - 29**
	Collaborating with our fact-checking partners in relation to emerging trends We have engaged with our fact-checking partners to understand and receive advance notice of emerging misinformation trends in order to ensure that our moderation teams are well-equipped to take action.	Since the beginning of the crisis, representatives from TikTok's misinformation policy and detection teams have met regularly with our external fact-checking partners, to receive advance warning of emerging misinformation narratives, facilitating proactive responses against high-harm trends.	**Number of videos removed because of policy guidelines, known misinformation trends and knowledge based repository with words related to RU/UA - 1292**

Measures taken to demonetise disinformation related to the crisis	**Prohibiting political advertising and monetization related to the war from our platform**	We do not allow political advertising from anyone, including from state media, on our platform. This has been <u>our position</u> since October 2019 and it has not altered since the beginning of the war in Ukraine. More specifically, and as outlined in our <u>advertising policy</u>, we do not permit ads for the promotion, reference, or opposition of: a candidate for public office, a current or former political leader, political party, or political organisation, content that advocates a stance (for or against) on a local, state, or federal issue of public importance in order to influence a political decision or outcome. As mentioned previously, our monetisation integrity content teams observed a sharp rise in attempts to post ads related to political/war content during the initial period of the conflict commencement. Our risk control methods ensured there was no corresponding increase in violating activity going live on the platform. Since the initial period of the conflict, these numbers have significantly dropped. Since the beginning of the war, we have taken a strict position not to enable associated monetization. This includes direct monetisation (e.g. the sale of merchandise associated with the war) as well as any reference to the war. Since March 2022, we have not allowed Ukrainian and Russian advertisers to target EU markets.	**N/A**
Measures taken to prevent malicious advertising	**Content moderation** We use a combination of automated and human moderation in order to identify content that breaches our ad policies.	We enforce our strict advertising policies with a combination of automated and human moderation, and have expert teams focused on investigating and responding to any attempts to circumvent our policies. Our Monetization Integrity department has moderation teams in multiple locations that speak Russian and Ukrainian. Given the range of potential policy violations that could be engaged, we are currently unable to provide metrics specific to this issue.	**N/A**
Measures taken in the context of the crisis to counter manipulative behaviours/TTPs	**Identifying and removing CIO networks** Our dedicated investigation teams and automated systems have been on high alert to identify threats related to CIO networks and have removed networks targeting discourse about the war in	As well as our I&A policies in our CGs which safeguard against harmful misinformation, our integrity policies also robustly prohibit deceptive behaviours such as CIO networks. We have specifically-trained teams which are on high alert to investigate and detect CIO networks on our platform. Between July and September 2022, we identified and removed 2 distinct networks targeting discourse about the war and 1,704 associated accounts, for coordinated efforts to influence public opinion	*NB: We are only able to provide the Q3 2022. We are building and testing data infrastructure that can provide information requested at a high level of fidelity. Additionally, CIO network investigations are highly resource heavy*

	Ukraine, in accordance with our I&A policies which prohibit deceptive behaviours.	and mislead users about their identities. We published this information voluntarily within our most recently published transparency report, here. We know that CIO will continue to evolve in response to our detection and networks may attempt to reestablish a presence on our platform, which is why we continually seek to strengthen our policies and enforcement actions in order to protect our community against new types of harmful misinformation and inauthentic behaviours. We have strengthened our efforts to detect suspected accounts impersonating high profile individuals/entities in the affected regions in order to avoid the spread of misinformation or any kind of misleading claim.	*requiring in depth analysis to ensure high confidence in proposed actions.* **Number of impersonation accounts related to RU/UA (prominent politicians) - 2**
Measures taken to support research into crisis related misinformation and disinformation	We have not been involved in any research initiatives in the context of the war to date, however, we are committed to engaging with researchers going forwards and hoping to pilot API testing soon.	We have not been involved in any research initiatives in the context of the war to date. However, we will soon be launching a pilot phase as part of our efforts to develop an API for researchers as we are committed to engaging with the research community and ensuring that our tools meet their needs going forwards.	**N/A**
Relevant changes to working practices to respond to the demands of the crisis situation and/or additional human resources procured for the mitigation of the crisis	We have continued to invest in localised moderation capabilities.	We have continued to invest within the region, including Russian and Ukrainian native language speakers for content moderation, and this investment has been significantly increased since the war began.	This is a dynamic situation, and we also leverage the capabilities of fact-checking partners, so numbers can fluctuate.

Covid-19 pandemic

Overview of the main threats observed, such as crisis related disinformation campaigns, spread of misinformation, coordinated manipulative behaviours, malicious use of advertising products, involvement of foreign state actors, etc.:

We continue to see a decrease across our metrics in conjunction with the lifting of restrictions in many countries in the EU and the return to "normal life". We also note that the number of Covid-19 vaccine tags are exceeding the Covid-19 notice tags, likely due to the fact that the vast majority of Covid-19 related content is now in relation to vaccines/vaccination.

We have not seen any evidence of covert influence operations (**CIO**) on our platform as regards Covid-19 disinformation in the EU. We also have not observed coordinated attempts to influence or sway public opinion on Covid-19 pandemic while also misleading individuals, our community, or our systems about an account's identity, approximate location, relationships, popularity, or purpose.

TikTok's advertising policy for Covid -19 related ads remains unchanged, with permitted ads generally coming from government authorities only. We are currently seeing very few attempts to post ads related to Covid-19 or Covid-19 vaccines.

Executive summary of the company's main strategies and actions taken to mitigate the identified threats and react to the crisis:

We are continually working hard to ensure that TikTok is a source of reliable and safe information and recognise the heightened risk and impact of misleading information during a time of crisis.

Following the outbreak of the Covid-19 pandemic, we introduced several new policies to our Integrity & Authenticity Policies (**I&A policies**) including our Medical Misinformation policy, in order to respond to trends we witnessed and protect our users. We employ a layered approach to misinformation detection, leveraging multiple overlapping strategies to ensure comprehensive and responsive coverage and as part of this we have specialised misinformation moderation teams working to identify and remove such false or misleading content, as well as accounts that spread misinformation. Since the beginning of the Covid-19 pandemic, we have removed hundreds of thousands of videos containing misinformation related to the pandemic. At TikTok, we invest considerable resources in proactively identifying violative content, and so are proud that the vast majority of these videos were removed by us before they were reported by users or other third parties. We note that during the pandemic we also made changes to our platform to ensure that users are able to easily report videos they believe contain misinformation specifically about Covid-19.

Working closely with our fact-checking partners has been a crucial part of our approach to removing harmful medical misinformation from our platform. In particular, collaborating with certain of our fact-checking partners to receive advance warning of emerging misinformation narratives (including those related to vaccines and false cures) has facilitated proactive responses against high-harm trends and has ensured that our moderation teams have up-to-date guidance.

Recognising the importance of proactive measures, which are aimed at improving our users' digital literacy and increasing the prominence of authoritative information, we created an in-app hub information hub dedicated to disseminating authoritative Covid-19 information from the WHO to our users. Since July 2020, the hub has been viewed more than 5 billion times globally. In January 2023, we made the decision to discontinue the in-app hub and have updated our dedicated Covid-19 notice tags and search intervention tools so that they direct users straight to authoritative, localised information from expert organisations. We are also proud to have partnered with a number of other experts in relation to the pandemic to share authoritative information with our users, including Team Halo.

In relation to ads, we developed our Covid-19 advertising policy, which prohibits ads which present Covid-19 in a distasteful manner e.g., manipulating consumers' fear or anxiety, or spreading harmful misinformation to push sales. We also promoted authoritative sources of information by providing free ad credits to health authorities, governments, and non-profits to increase vaccine adoption, debunk vaccine myths and promote trustworthy Covid-19 resources.

Best practices identified for future crisis situations:

- Working closely with our fact-checking partners and other experts to receive advance warning of emerging misinformation narratives. For TikTok, this has facilitated proactive responses against high-harm trends and has ensured that our moderation teams have up-to-date guidance.
- Dedicating resources to increasing the number of in-app measures that show users additional context on certain content or redirect them to authoritative information.
- Rolling out media literacy campaigns in partnership with trusted organisations, focusing on addressing disinformation.

Future measures planned within the next 6 months:

- Driving forwards our ambitious plans to further expand our fact-checking network, by adding new fact-checking partners and increasing the number of languages our current fact-checking partners provide support in. Over the next year, we will be focusing on setting up fact-checking partnerships in Portugal, Denmark, Greece and Belgium as well as other EU countries. As part of these plans, we will be expanding our misinformation moderation teams.
- Continuing to work with our fact-checking partners to identify specific disinformation trends in countries and develop tailored, localised media literacy campaigns to tackle those trends.
- Redirecting our users to authoritative, localised information from expert organisations (such as local public health sites or, where local health sites are not available, the WHO) when they interact with our dedicated Covid-19 notice tags and search intervention tools.

Changes in Policy Framework		
	Policies	**Rationale**
Policies newly introduced for addressing the crisis	Since the beginning of the reporting period, we have been able to rely on our existing, robust I&A policies as well as our dedicated advertising policies and as such have not needed to introduce new policies for the purposes of addressing the crisis.	In the context of the Covid-19 pandemic, we have been able to rely on our existing I&A policies which are our first line of defence in combating harmful misinformation and deceptive behaviours on our platform. These guidelines make clear to our users what content we remove or make ineligible for the For You feed when they pose a risk of harm to our users and their communities. Following the outbreak of the Covid-19 pandemic, we introduced several new policies to our Community Guidelines (**CGs**) I&A policies, in order to respond to trends we witnessed and protect our users. The policies we introduced, as well as examples as to the basis on which they have been enforced in relation to the crisis are as follows: • **Medical misinformation that can cause harm to an individual's physical health.** Under this policy, we have been able to remove false claims about coronavirus, false cures and vaccine misinformation that could cause imminent harm to a user's health, or could cause harm to public safety by discouraging people from seeking helpful medical treatment.

		• **Misinformation related to emergencies that induces panic.** Under this policy, we have been able to remove content that is likely to stoke panic and consequently result in real world harm (e.g. content encouraging destruction of 5G towers by claiming they are spreading Covid-19). • **Misinformation that incites hate or prejudice.** In accordance with this community guideline and our hate speech policy, we have been able to remove content that suggests a certain race, ethnicity, gender or any member of a protected group is more susceptible to having and / or spreading Covid-19. We have also supported our moderation teams with detailed policy guidance and direction when moderating on crisis related misinformation providing clarifications, updates as required. We have specialist teams within our Trust and Safety department dedicated to the policy issue of Integrity and Authenticity, including within the areas of product and policy. Our experienced subject matter experts on Integrity and Authenticity continually keep these policies under review and collaborate with external partners and experts when understanding whether updates are required. In an ads context, we have been able to rely on our Covid-19 advertising policy which prohibits ads which present Covid-19 in a distasteful manner, which includes spreading harmful misinformation to push sales.
Policies adapted for addressing the crisis	Since the beginning of the reporting period, we have not adapted our Terms of Service or CGs for addressing the crisis, as we have been able to rely on the I&A policies as well as our dedicated advertising policies as an effective basis for tackling content related to the pandemic.	See rationale above.
Actions to mitigate the crisis impact on the service		
Type of mitigation	**Intervention or action (short summary)**	**Intervention or action (explanation and implementation)**	**Impact metrics** ***NB:*** *the impact metrics we are sharing within this report are based on proxies and, as such, should be regarded as an approximation. Our reporting period for the*

			purpose of these metrics is October - December 2022.
Actions taken against dis- and misinformation content (for example deamplification, labelling, removal etc.)	**Removing harmful misinformation from our platform** Since the beginning of the Covid-19 pandemic, we have removed hundreds of thousands of videos containing misinformation related to the pandemic. At TikTok, we invest considerable resources in proactively identifying violative content, and so are proud that the vast majority of these videos were removed by us before they were reported by users or other third parties.	Our I&A policies prohibit content which contains false or misleading claims that may cause harm to individuals, our community or the larger public, regardless of intent. In light of the Covid-19 pandemic, we updated these policies to cover medical misinformation (which includes false medical advice about Covid-19 and Covid-19 vaccines), misinformation likely to cause societal panic and real-world harm and hate speech (which would include any suggestion that a group of people with certain characteristics are more likely to have or spread Covid-19). We employ a layered approach to misinformation detection, leveraging multiple overlapping strategies to ensure comprehensive and responsive coverage and as part of this we have specialised misinformation moderation teams working to identify and remove such false or misleading content, as well as accounts that spread misinformation. Throughout the pandemic, we have ensured that our moderators are updated on new and continuing false Covid-19 narratives, including those related to vaccines and false cures. At TikTok, we place considerable emphasis on proactive content moderation and so are proud that the vast majority of videos removed for violating our harmful misinformation policies were identified and removed by us (before being reported to us by users or other third parties). Our community members are able to report misinformation to us from our app and during the pandemic, we made changes to our platform to ensure that users are able to easily report videos they believe contain misinformation specifically about Covid-19. If a user comes across content on TikTok that appears to contain misleading information about coronavirus, they can report the content, select "Misleading information" and then "Covid-19 misinformation" as the reason for their report. Each report is reviewed promptly by a dedicated team and appropriate action is then taken.	**Number of videos removed because of violation of misinformation policy with a proxy related to Covid19 and Covid-19 Vaccine - 1802** **Number of proactive removals of videos removed because of violation of misinformation policy with a proxy related to Covid 19 and Covid Vaccine - 1557**
	Deamplifying unverified content by applying our unverified content label Where our specialised misinformation moderators or fact-checking partners	Verifying certain information during dynamic and fast moving events such as a global health pandemic can be challenging and our moderators and fact-checkers cannot always conclusively determine whether content is indeed harmful misinformation, in violation of our CGs.	*NB: As mentioned above, we continue to see a decrease across our metrics in conjunction with*

	determine that content is not able to be verified at the given time (which is common during an unfolding crisis), we apply our unverified content label to the content to encourage users to consider the reliability or source of the content. The application of the label will also result in the content becoming ineligible for recommendation in order to limit the spread of misleading information.	Therefore, in order to minimise risk, where our fact-checkers or our specialised moderators do not have enough information to verify content which may potentially be misleading, we apply our unverified content label to inform users the content has been reviewed but cannot be conclusively validated. The goal is to raise users' awareness about the credibility of the content and to reduce sharing (see screenshots here). Our unverified content label is available in 21 EU official languages (plus, for EEA users, Norwegian and, as the spoken language of Liechtenstein, German) Where the banner is applied, the content will also become ineligible for recommendation into anyone's For You feed to limit the spread of information relating to unfolding events where details are still developing and which may potentially be misleading. We also prompt users if a video has been flagged as containing unsubstantiated content. If a user then tries to share the video, they will be asked to pause and consider their next move before they choose to 'cancel' or 'share anyway'. We are working to improve the accuracy of the data metrics in respect of not recommend moderation actions and will include these in future reports once that work is complete.	*the lifting of restrictions in many countries in the EU and the return to "normal life".* **Number of videos tagged with the unverified content label - 29** **Share cancel rate for unverified content label - 40%**
	Labelling content regarding Covid-19 and vaccines We apply notice tags to Covid-19 and vaccine related videos shared on our platform. The notice tags direct users to authoritative information in order to reduce the risk of disinformation spreading.	Using automated models, we identify and review content related to Covid-19 and vaccines and apply video notice tags. The notice tag appears as a banner on the bottom of the video and redirects users to verifiable, authoritative sources of information within our in-app Covid-19 hub, in order to mitigate the risk of disinformation spreading on our platform. We have set out further information regarding our Covid-19 hub below.	**Impressions to our Covid-19 and vaccine notice tags - 1429860795** **Clicks to our Covid-19 and vaccine notice tags - 1,563,872** **Click through rate of our Covid-19 and vaccine notice tags - 0.11%**
Promotion of authoritative information, including via	**Used our in-app intervention tools to share authoritative Covid-19**	We are proud to have been one of the first content platforms to have worked with public health experts to make authoritative	**Views of the Covid 19**

recommender systems and products and features such as banners and panels	**information with our users** We were one of the first content sharing platforms to create an in-app hub information hub dedicated to disseminating authoritative Covid-19 information to our users. Since July 2020, the hub has been viewed more than 5 billion times globally. We have recently taken the decision to discontinue the hub and update our dedicated Covid-19 notice tags and search intervention tools such that they direct users straight to localised information from public health authorities (or the WHO).	information about Covid-19 and Covid-19 vaccines available in a dedicated in-app Covid-19 information hub. We used our search intervention tools effectively in order to ensure that as many users as possible were directed towards the hub. Most notably, we ensured that the hub was pinned to our search banner when users searched for Covid-19 related content or hashtags within the app. In addition, and as mentioned above, our notice tags directed users to the hub. Within the hub, our community were able to find answers to common questions about the virus and vaccines from the WHO and local health authorities, as well as tips on staying safe. In January 2023, we made the decision to discontinue the in-app hub and have updated our dedicated Covid-19 notice tags and search intervention tools so that they direct users straight to authoritative, localised information from expert organisations (such as local public health sites or, where local health sites are not available, the WHO).	**Information hub - 559406** **Impressions to our Covid-19 <u>and</u> vaccine Search Intervention - 1668679** **Clicks to our Covid-19 <u>and</u> vaccine Search Intervention - 15,661** **Clicks through rate of our Covid-19 <u>and</u> vaccine Search Intervention - 0.96%** **Impressions to our Covid-19 <u>and</u> vaccine notice tags - 1429860795** **Clicks to our Covid-19 <u>and</u> vaccine notice tags - 1,563,872** **Clicks through rate of our Covid-19 <u>and</u> vaccine notice tags - 0.11%**
	Creating our Covid-19 Centre We created a dedicated Covid-19 centre within our Safety Center in order to clearly set out for our users certain measures we have taken specifically in response to the crisis in order to help them stay safe and informed. See the centre **here**.	In 2021, we launched a redesigned [Safety Center](#) (all users received an in-box notification about this). Within the centre, we created a designated page regarding [Covid-19](#). The page sets out some of the steps we have taken to help our community stay safe and informed in the context of the Covid-19 pandemic. For example, the centre contains information in relation to some of the ways in which we combat medical misinformation as well as certain organisations who use our platform to share trusted information (for example, highlighting that the WHO was serving as a valuable resource to our community by using their account to create videos highlighting key tips and facts).	**Page views and User views to COVID Centre Page - 444883**

	Partnering with [Team Halo](#) to share authoritative vaccine information We are proud to have collaborated with Team Halo in order to allow top scientists and clinicians to update users on vaccines.	Our partnership with [Team Halo](#) continues as a means to update users on vaccines. Through our Team Halo partnership, scientists from all over the world continue to post video updates on TikTok to show the world the progress being made on the vaccine. This team of public health experts answer all kinds of questions from our community, from what steps go into developing a vaccine to how they test them for safety.	
Cooperation with independent fact-checkers in the crisis context, including coverage in the EU	**Ensuring fact-checking coverage and collaborating with our fact-checking partners in relation to emerging trends** Our fact checking efforts cover all major European languages (including 10 official European languages as well as a number of other languages which affect European users). We have engaged with our fact-checking partners to understand and receive advance notice of emerging misinformation trends in order to ensure that our moderation teams are well-equipped to take action.	We partnered with fact checkers throughout the EU in relation to Covid-19 misinformation, including AFP, Facta, Logically, Lead Stories, Newtral, Science Feedback, Teyit and DPA. We have engaged with our fact-checking partners to understand and receive advance notice of emerging misinformation trends in order to ensure that our moderation teams are well-equipped to take action. Since the beginning of the pandemic, representatives from TikTok's misinformation policy and detection teams have met regularly with our external fact-checking partners, to receive advance warning of emerging misinformation narratives, facilitating proactive responses against high-harm trends. Note that we are continuing to expand our fact-checking program both in terms of the EEA countries in which we have fact-checking partners (and thus onboarding new partners) and the scaling up the volume of claims we fact-check and content our current partners are reviewing. As mentioned above, over the next year, we are focusing on setting up fact-checking partnerships in Portugal, Denmark, Greece and Belgium as well as other EU countries. As we build momentum with our fact checking partners, our metrics will reflect this progress.	**Number of fact checked videos related to Covid-19 - 57** **Number of videos removed as a result of a fact checking assessment (with words such as Covid-19 and Covid-19 Vaccine) - 30**
Measures taken to demonetise disinformation related to the crisis	**Prohibiting advertising that advocates against vaccinations and providing free advertising space to health authorities** We developed our Covid-19 advertising policy to prohibit ads which present Covid-19 in a distasteful manner (including spreading harmful	We have developed specific advertising policies that focus on certain topics where the risk of disinformation may be higher, with a view to taking a targeted approach to defund the spread of disinformation. Our [Covid-19 advertising policy](#) prohibits ads which present Covid-19 in a distasteful manner e.g., manipulating consumers' fear or anxiety, or spreading harmful misinformation to push sales.	**Ads enforcement actions volume by Covid-19 policy - 6**

	misinformation to push sales).	As well as ensuring ads relating to Covid-19 do not spread harmful misinformation, we also promoted authoritative sources of information. We provided free ad credits to health authorities, governments, and non-profits to increase vaccine adoption, debunk vaccine myths and promote trustworthy Covid-19 resources.	
Measures taken to prevent malicious advertising	**Content moderation** We use a combination of automated and human moderation in order to identify content that breaches our policies.	We enforce our strict advertising policies with a combination of automated and human moderation, and have expert teams focused on investigating and responding to any attempts to circumvent our policies. Given the range of potential policy violations that could be engaged, we are currently unable to provide metrics specific to this issue.	**N/A**
Measures taken in the context of the crisis to counter manipulative behaviours/TTPs	We have remained alert to the risk of manipulative behaviours and TTPs in the context of Covid-19, however, we have not observed any such activity to date.	We work with industry-leading threat detection platforms to identify networks and suspicious activity and remain vigilant to such threats. However, we note that we have not seen any evidence of CIO networks on our platform as regards Covid-19 disinformation in the EU. We also have not observed coordinated attempts to influence or sway public opinion on Covid-19 pandemic while also misleading individuals, our community, or our systems about an account's identity, approximate location, relationships, popularity, or purpose.	**N/A**
Measures taken to support research into crisis related misinformation and disinformation	We have not been involved in any research initiatives in the context of the Covid-19 pandemic to date, however, we are committed to engaging with researchers going forwards and hoping to pilot API testing soon.	We have not been involved in any research initiatives in the context of the Covid-19 pandemic to date. However, we will soon be launching a pilot phase as part of our efforts to develop an API for researchers as we are committed to engaging with the research community and ensuring that our tools meet their needs going forwards.	**N/A**
Relevant changes to working practices to respond to the demands of the crisis situation and/or additional human resources procured for the mitigation of the crisis	We have continued to invest in localised moderation capabilities.	We continue to invest in content moderation, and this investment has also covered our needs during the Covid-19 pandemic.	This has been a dynamic situation throughout the pandemic and we have also leveraged the capabilities of expert fact-checkers. As such, data has fluctuated.