2022 Annual Transparency Report

TikTok

Australian Code of Practice on Disinformation and Misinformation

January 2022 – December 2022

Table of Contents

Introduction

Approach

As a leading destination for short-form mobile video, TikTok is committed to nurturing creativity in a safe, supportive and authentic environment. We recognise the impact misinformation can have in eroding trust in public health, electoral processes, facts, and science - and we are committed to being part of the solution. We treat misinformation with the utmost seriousness and take a multi-pronged approach to stopping it from spreading. This approach includes proactive detection of violations, enhanced user education, accredited partnerships and transparent reporting to help TikTok combat these and other integrity risks at scale.

To promote an authentic user environment, the "Integrity and Authenticity" (I&A) policies within our Community Guidelines[1] (CGs) prohibit misinformation, impersonation, and coordinated or manipulated misleading content. Violative videos are removed from the platform, and these efforts are detailed in TikTok's quarterly Community Guidelines Enforcement Reports. **Fig. 1** shows the removal of content violating our I&A policies by quarter, outlining the proportion of I&A violations against all removed content, as well as statistics on removal efficiency. "Proactive removal" in this context refers to policy enforcement before it is reported by users.

Quarter (2022)	% from all CG violations	Proactive removal	Removal before content receives any views	Removal within 24 hours of content posted to platform	
January - March	0.6%	83.6%	60.8%	71.9%	
April - June	0.7%	89.1%	74.7%	83.95	
July - September	0.7%	92%	77.2%	83.5%	
October - December	0.9%	91.5%	69.5%	71.3%	

Fig.1: *Removal of I&A Violations in 2022*

Over 2022, TikTok undertook a number of projects to strengthen authenticity assurance globally as a response to local Australian events and unfolding global issues. Among others, these included projects undertaken in preparation for and in response to the Australian federal election, the Victorian state election, the COVID-19 pandemic, Russia's

---

[1] Community Guidelines are applicable as of year 2022.

![TikTok logo]

invasion of Ukraine and the Turkey-Syria earthquake. Key initiatives undertaken by TikTok to support long-term risk mitigation include:

- investment in machine learning models to ensure extensive coverage of nuanced misinformation threats;
- discovery of inauthentic visual and audio trends to combat manipulated, edited and deepfake content;
- a database of fact-checked claims enabling human moderators to accurately identify misinformation content;
- an anti-misinformation program in partnership with accredited fact-checkers to assess inauthentic narratives on third party platforms and restrict them from TikTok.

Over 2022, these projects drove continued, measurable improvements in our detection of, and response to, misleading content. **Fig. 2** shows the quarterly removal rates of content containing harmful misinformation as a proportion of total I&A policy violations and associated statistics on enforcement efficiency. These results show that our proactive misinformation removal rates improved over the course of 2022 from 66.1% to 89.1%, with significantly more content detected by TikTok before it is reported by users. These figures also show an increased rate of removal of misinformation prior to user discovery from 37.6% in Q1 to 69.8% in Q4, underscoring both the effectiveness of strategies undertaken throughout 2022 and the further scope for future development.

Quarter (2022)	% within I&A policy	Proactive removal	Removal before content receives any views	Removal within 24 hours of content posted to platform
January - March	33.3%	66.1%	37.6%	50.6%
April - June	28.3%	70.0%	39.1%	57.5%
July - September	30.6%	78.8%	55.3%	59.6%
October - December	61.1%	89.1%	69.8%	59.9%

Fig.2: Removal of Harmful Misinformation videos in 2022

Notable Highlights

Over 2022, TikTok introduced multiple initiatives - both globally and specific to Australia - to fight misinformation and provide a safe user experience. These included:

- Supporting the Australian federal election through active collaboration with the Australian Electoral Commission.

- [Launching the 2022 Australian Election Guide](#) to promote authentic information and encourage voting.

- Providing public service announcements (PSAs) for hashtags related to the Australian Federal Election, accumulating ~90,000 page views.

- Establishing [fact-checking partnerships](#) with [Australian Associated Press](#) (AAP) and [Agence France-Presse](#) (both of which are [IFCN-certified](#)) to prevent the spread of misleading information.

- Supporting the Victorian state election through [PSAs](#) for hashtags and search terms, reminding users of our Community Guidelines and directing users to the Victorian Electoral Commission's website.

- Partnering with Melanoma Institute Australia to launch the "[Tanning. That's Cooked](#)" education campaign, targeting Australians aged 20 to 39 to raise awareness of the risks and hazards of tanning.

- Engaging in strategic research to understand perceptions of misinformation by the Australian user community and the confidence of users in identifying and reporting such content.

- Delivering a digital media literacy campaign to top creators from New Zealand, including educational materials on sourcing authentic information, critically approaching new data and creating reliable, informative content.

- Developing and releasing new features enabling users to learn why particular videos are recommended to them on their "For You" feed, providing enhanced transparency of TikTok's recommendation system.

- Publishing PSAs referring users to health information centres, electoral commissions, current affairs and mental health hotlines to support vulnerable users.

TikTok's Commitments under the Code

As a dedicated signatory, TikTok **opts in to all Objectives and Outcomes** under the Australian Code of Practice on Disinformation and Misinformation.

Objective 1: Provide safeguards against Harms that may arise from Disinformation and Misinformation
• **Outcome 1a:** Signatories contribute to reducing the risk of Harms that may arise from the propagation of Disinformation and Misinformation on digital platforms by adopting a range of scalable measures
• **Outcome 1b:** Users will be informed about the types of behaviours and types of content that will be prohibited and/or managed by Signatories under this Code
• **Outcome 1c:** Users can report content or behaviours to Signatories that violate their policies under section 5.10 through publicly available and accessible reporting tools.
• **Outcome 1d:** Users will be able to access general information about Signatories' actions in response to reports made under 5.11.
• **Outcome 1e:** Users will be able to access general information about Signatories' use of recommender systems and have options relating to content suggested by recommender systems.
Objective 2: Disrupt advertising and monetisation incentives for Disinformation and Misinformation.
• **Outcome 2:** Advertising and/or monetisation incentives for Disinformation and Misinformation are reduced.
Objective 3: Work to ensure the integrity and security of services and products delivered by digital platforms
• **Outcome 3:** The risk that Inauthentic User Behaviours undermine the integrity and security of services and products is reduced.
Objective 4: Empower consumers to make better informed choices of digital content.
• **Outcome 4:** Users are enabled to make more informed choices about the source of news and factual content accessed via digital platforms and are better equipped to identify Misinformation.
Objective 5: Improve public awareness of the source of Political Advertising carried on digital platforms.
• **Outcome 5:** Users are better informed about the source of Political Advertising.
Objective 6: Strengthen public understanding of Disinformation and Misinformation through support of strategic research.
• **Outcome 6:** Signatories support the efforts of independent researchers to improve public understanding of Disinformation and Misinformation.

Objective 7: Signatories publicize the measures they take to combat Disinformation and Misinformation.

- **Outcome 7:** The public can access information about the measures Signatories have taken to combat Disinformation and Misinformation.

The following sections of the report outline the specific measures, policies and projects undertaken to promote authenticity and address misinformation on TikTok.

Reporting against 2022 Commitments

Objective 1: Provide safeguards against Harms that may arise from Disinformation and Misinformation

Outcome 1a: Reducing harm by adopting scalable measures

TikTok has adopted a range of clear, scalable approaches to provide safeguards against harms that might arise from both mis- and disinformation. These include:

- information tags for content related to specific subject matter;
- policy enforcement through the removal of content that violates our Integrity and Authenticity policies; and
- in specific, warranted circumstances - account or device bans.

COVID-19 and vaccination information tags

Throughout 2022, we continued to refine and enforce our I&A policies in relation to COVID-19 and vaccine-related information. A major scalable intervention to tackle mis- and disinformation was our addition of information tags in relation to COVID-19 and vaccine-related content. These in-app disclaimers for specific videos seek to educate users and direct them to authoritative sources of information. **Fig. 3** shows the number of Australian TikTok videos tagged with COVID-19 information notices over 2021 and 2022.
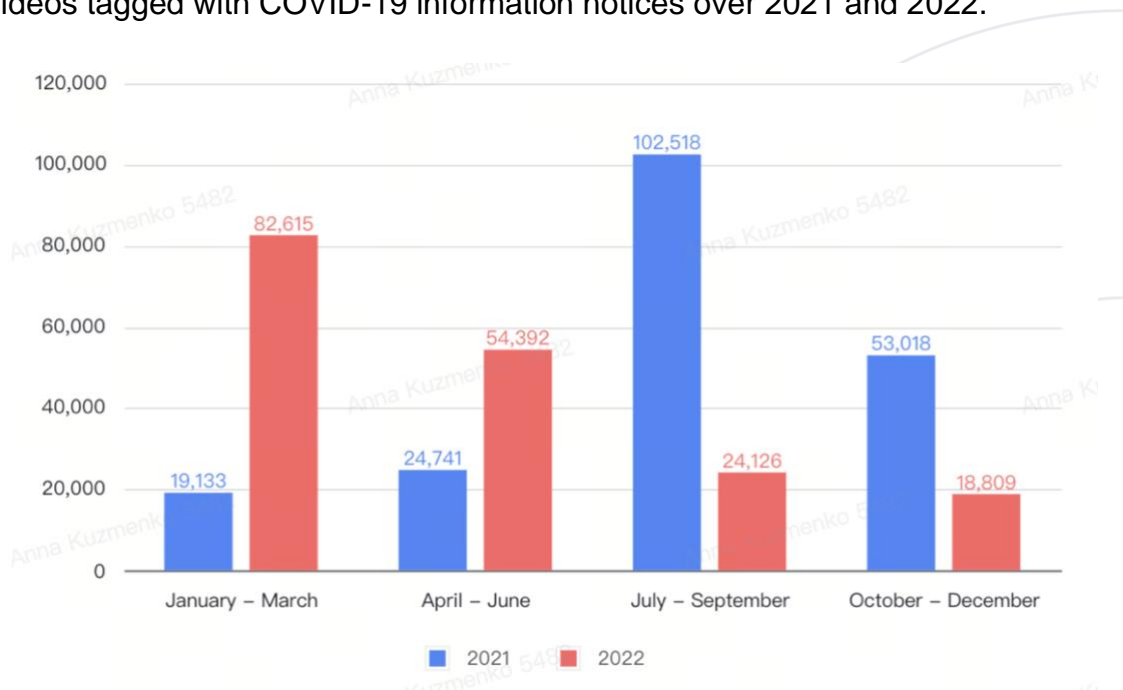

Fig.3: *Number of AU videos tagged with COVID-19 information notice*

In total, 199,410 videos were tagged with COVID-19 notice in 2021 and 179,942 in 2022. The data reveals a decrease in content notice tagging over 2022. Both external and platform-specific factors might account for this trend, including:

- a decrease in local content being identified as warranting information tags in 2022;
- less frequent discussion among Australian users as COVID restrictions were lifted and vaccine requirements eased over the course of 2021-22; and
- the effectiveness of account-level measures undertaken to reduce the ability of bad actors to spread medical misinformation related to COVID-19 and vaccines on our platform.

Content removal

We continue to remove, or make harder to find, content that violates our Community Guidelines in relation to I&A policies. Specifically, in relation to dangerous misinformation, election misinformation and medical misinformation, we removed a cumulative total of 15,702 videos in Australia in 2022.
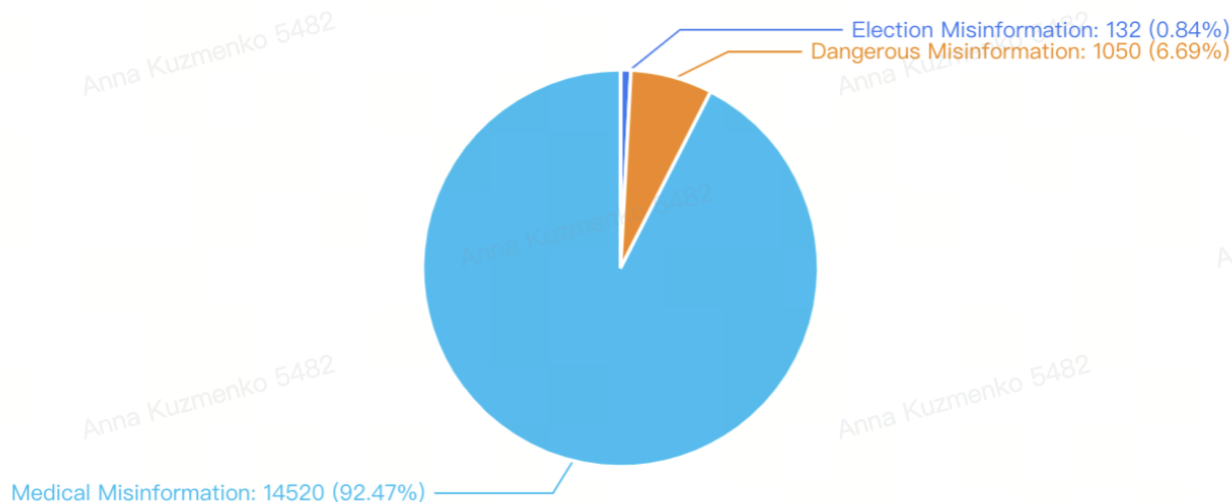
Fig.4: Distribution of removed videos violating misinformation policies in AU in 2022

Election-related policy violations remain relatively rare in Australia and are generally time-limited to the period close to elections themselves. This signals the robustness of both the systems we employ and the additional supporting measures TikTok adopts to ensure election integrity, such as dedicated escalation channels and third-party fact-checking. **Fig. 5** shows the monthly number of videos removed for violating our misinformation policies over 2022 in Australia. Compared to 2021, when 12,582 medical misinformation videos

were removed, TikTok removed 14,520 videos violating medical misinformation policy in AU in 2022.

Fig.5: *Number of monthly removed videos violating misinformation policies in AU in 2022*

Strengthening our resources to address misinformation and disinformation

We continue to build on our significant investments in the capacity of our teams, processes and technology to counter the spread of mis- and disinformation and address threats that may arise from known risks, emerging violative trends, or which reflect country-specific incidents. Our Trust & Safety teams number over 40,000 people globally, including subjec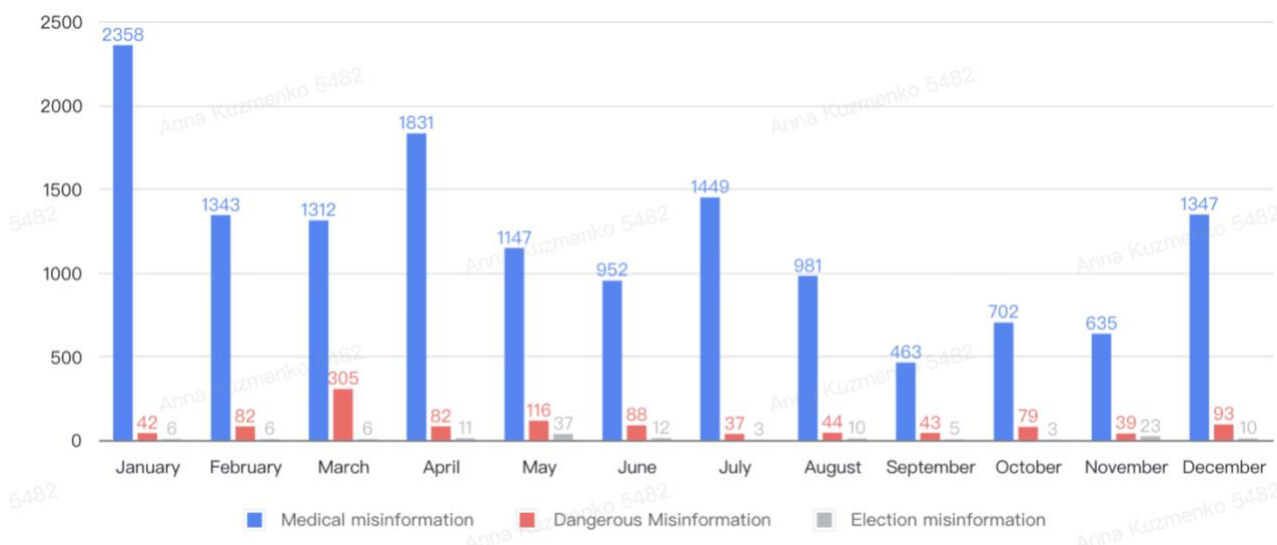t matter experts tasked with developing, implementing and reviewing TikTok's "Integrity and Authenticity" policies. We continue to refine and adapt our activities and processes in real-time to identify and combat inauthentic content. Recently, this has included implementing additional measures to detect and address violative livestreamed content.

Our 24/7 moderation capacity is supported by our third-party fact-checking partners who review and verify flagged content. These partnerships help minimise risks of mismoderation and ensure that moderation decisions are based on independent assessments by accredited third-party experts. This is critical for issues involving public health or civic processes, where the consequences of disseminating misinformation are particularly acute. In all cases where information is found to be false or deceiving, we take immediate measures to remove the content in line with our Community Guidelines. In 2022, we undertook a number of additional measures to counter misinformation, including:

- improved detection of known misleading audio and imagery to reduce manipulated content spread;

9

- a database of previously fact-checked claims to help misinformation moderators make swift and accurate decisions; and
- a proactive detection program for our fact-checkers to flag new and evolving trends they discover across the internet. This allows TikTok to retrieve and remove violative content from our platform. Since starting this program last quarter, we have identified 33 new misinformation claims, resulting in the removal of 58,000 videos from the platform globally (as of April 2023).

Outcome 1b: Inform users about what content is targeted

TikTok's Community Guidelines define a set of norms and common code of conduct for TikTok. The guidelines are available to users within the app and on our website. They detail what is and is not allowed on the platform, ensuring TikTok remains a welcoming space for creativity and entertainment. The relevant Community Guidelines regarding harmful misinformation are set out within our "Integrity and Authenticity" policies:

Do not post, upload, stream, or share:
- Misinformation that incites hate or prejudice;
- Misinformation related to emergencies that induces panic;
- Medical misinformation that can cause harm to an individual's physical health;
- Content that misleads community members about elections or other civic processes
- Conspiratorial content including content that attacks a specific person or a protected group, includes a violent call to action, or denies a violent or tragic event occurred;
- Digital Forgeries (Synthetic Media or Manipulated Media) that mislead users by distorting the truth of events and cause significant harm to the subject of the video, other persons, or society.

Do not:
- Engage in coordinated inauthentic behavior such as the use of multiple accounts to exert influence and sway public opinion while misleading individuals, our community, or our systems about the account's identity, location, relationships, popularity, or purpose.

In developing our policy positions, TikTok listens to the community and actively consults with NGOs, regulators, academics and other experts. We also rely on the advice of the TikTok Asia Pacific Safety Advisory Council, formed in 2020, to formally bring together a group of independent online safety experts to advise on issues related to trust and safety.

In February 2023, we made changes to our account enforcement system which now results in users being provided with clear reasons for account penalties they accrue, including

10

those associated with violations of the "Integrity and Authenticity" policy titles. These changes acknowledge that the previous system disproportionately impacted creators who rarely and unknowingly violated policies, while potentially being less efficient at deterring repeat offenders. Internal analysis showed that repeat violators tend to follow a pattern – almost 90% violate guidelines using the same feature consistently (video visuals, comments, etc.), and over 75% repeatedly violate the same policy category. To address this, we [updated our enforcement system](#) to better support our creator community and remove repeat offenders from the platform.

We also introduced several [feature updates](#) in February 2023 to provide users with clearer insights into their account status and reports. These include an "Account Status" tab, where creators can easily view the standing of their account, as well as a "Reports Records" page which displays the status of all their reports.
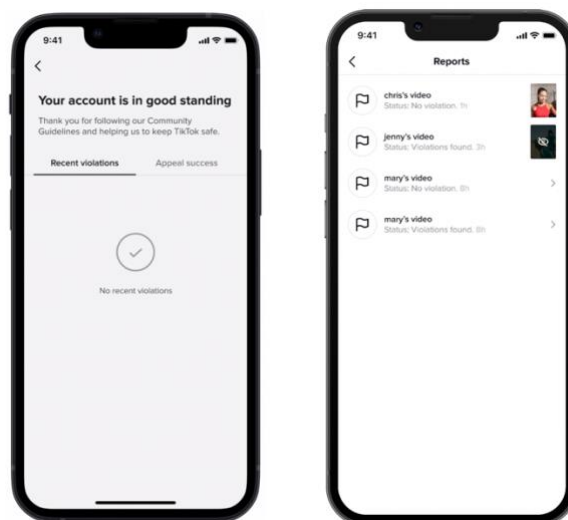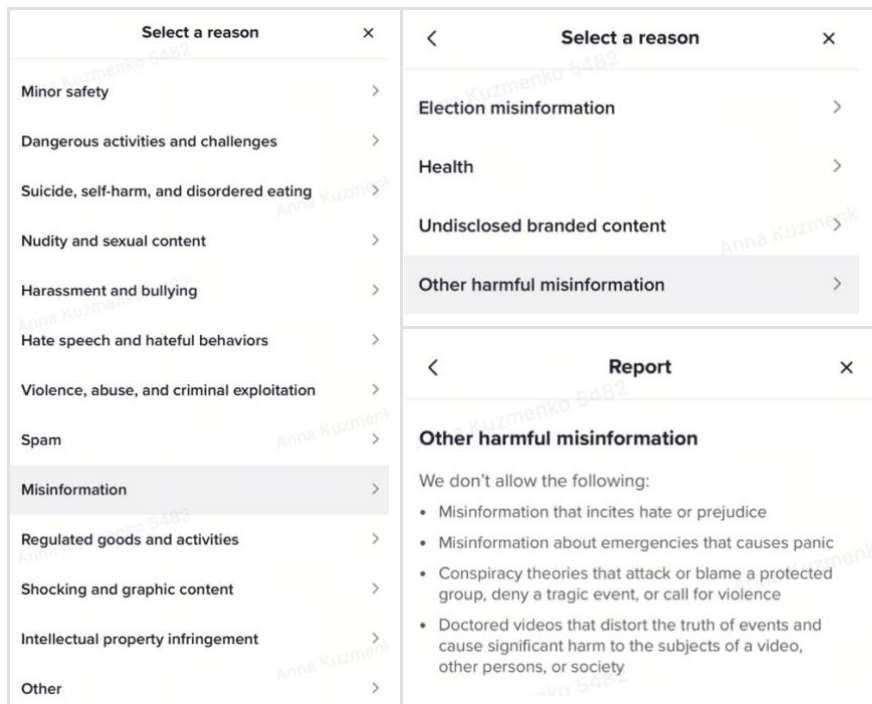
Fig.6: "Account Status" tab and "Reports Records" page

These new tools add to the notifications creators already receive when a policy violation is recorded, and support creators' ability to appeal enforcements and have strikes removed if their appeal is successful. We also notify creators if their account is at risk of a permanent ban. This reinforces TikTok's commitment to transparently informing users about content enforcement decisions that affect them.

Outcome 1c: Users can easily report offending content

TikTok users can [report content](#) they believe violates our [Community Guidelines](#) at any time within the app. 'Misinformation' is a distinct reporting category within our platform and appears as a prominent reporting option when users attempt to report content. Users are

11

![TikTok logo]

also prompted to specify the kind of misinformation they are reporting, as is shown in **Fig. 7** below.

Fig. 7: *User in-app reporting interface*

In addition to reporting misinformation in video content, users can also report misinformation across other features of the platform. Users can report comments on videos, direct messages they receive from other users, accounts, sounds, hashtags and auto-suggestions generated when they search for something on TikTok. Users can also report LIVE videos and comments on livestreams if they encounter content that violates our Community Guidelines.

Non-users can submit a report to us about content on our platform using the out-of-app reporting form. We have also established systems (including the Community Partner Channel) that enable direct reports of potential misinformation for our immediate review and action. The TikTok Community Partner Channel provides additional options for users to report content via trusted community organisations, who connect with community clients by sharing expertise and support on specific issues.

We also publish instructions for our publicly accessible reporting tools on our website.

Outcome 1d: Information about reported content available

Information about TikTok's approach to violations of our Community Guidelines, which include violations of our misinformation policies, is available on our website.

12

All creators are immediately notified when their content is found to have violated our Community Guidelines, and creators can view detailed information about these violations through the in-app features and account enforcement system detailed above. These features aim to enhance user awareness of our Community Guidelines, reduce ambiguity about permitted content and strengthen platform safety.

Our Transparency Centre is a one-stop shop to learn about how TikTok moderates and recommends content, develops products and safeguards user information.

Periodic reports published on the Transparency Centre comprise:

- Community Guidelines Enforcement Reports: providing quarterly insight into the actions we take to enforce our Community Guidelines and Terms of Service;
- Information Requests Reports: providing bi-annual data on user information requests from governments and law enforcement agencies and details of our response;
- Government Removal Requests Reports: providing bi-annual data on requests from government agencies to restrict content and any actions we take as a result; and
- Intellectual Property Removal Requests Reports: providing bi-annual data on copyright and trademark content takedown requests and our responses.

This data is available for download from our Transparency Centre in machine-readable formats, and can also be displayed in interactive charts and graphs to better illustrate these datasets. The reports are published in a range of languages.

Our latest Community Guidelines Enforcement Report shows that video removals pursuant to our I&A policies accounted for 0.9% of all videos removed from our platform globally in Q4 2022. Notably, 91.5% of this content was removed proactively by TikTok.

Outcome 1e: Users can access general information about Signatories' use of recommender systems and have options relating to content suggested by recommender systems.

TikTok publicly discloses information about our recommendation system through accessible explanatory material, and we have published a dedicated two-part Recommendation System "Q&A" for our users. This includes information on how and why videos appear in users' "For You" feeds, as well as details on video performance (e.g. number of views).

We also provide guidelines on recommendation systems and their navigation to our users (for instance, via blog posts). In March 2023, we introduced an option for users to refresh their "For You" feeds. When enabled, this feature allows users to view content on their "For You" feed as if they have just signed up for TikTok. Our recommendation system will then begin to surface more content based on new interactions. This provides users with more

13

flexibility in circumstances where there existing "For You" feeds may not feel relevant anymore, or [provide enough topical variety.](#)

Objective 2: Disrupt advertising and monetisation incentives for disinformation.

Outcome 2: Advertising and/or monetisation incentives for Disinformation and Misinformation are reduced.

As TikTok grows, we continue to maintain strong platform control by [strengthening our advertising policies](#). We do not allow the monetisation of government-owned accounts or political advertising, with the exception of cause-based advertising and information notices from non-profit or governmental organisations in collaboration with TikTok Sales Representatives. Our advertising policies also contain strict prohibitions on ads that contain deceptive or misleading claims, or which attempt to exploit or profiteer from sensitive events or subjects, such as the COVID-19 pandemic.

Global Alliance for Responsible Media (GARM) membership

As a member of [GARM](#), we collaborated with advertisers, agencies and other platforms in [GARM's Standards and Definitions Working Group](#) to develop a definition of misinformation to be included in the Global Brand Safety Floor and Suitability Framework. In June 2022, GARM published an [update to their framework](#) with the following definition of "misinformation":

- **Misinformation** is defined as the presence of verifiably false or willfully misleading content that is directly connected to user or societal harm.

This definition complements [current efforts](#) of the association to shift how brands, media agencies and ad tech partners operate, and the [framework](#) provides a process to demonetize misleading advertisements. As an active member of the GARM Working Group, TikTok is committed to contributing to further improvements to the framework's effectiveness.

14

Objective 3: Work to ensure the integrity and security of services and products delivered by digital platforms.

Outcome 3: The risk that Inauthentic User Behaviours undermine the integrity and security of services and products is reduced.

TikTok is committed to preventing, detecting and deterring influence operations on our platform, including by removing inauthentic accounts and tackling fake account engagement. Our integrity policies inform our approach to removing content and accounts involving spam or fake engagement to reduce the spread of misinformation on our platform. While both technology and safety professionals work together to enforce these policies, the challenge with misinformation and disinformation is unique as it typically requires greater contextual understanding and fact-checking than other areas of content policy enforcement. We currently have 15 fact-checking partners across the globe, working in over 40 languages to independently assess the veracity of claims made on our platform. All of our partners are verified signatories of the International Fact-Checking Network's code of principles. If fact-checkers confirm that content is false, it is removed from the platform or made ineligible for recommendation in "For You" feeds. Additionally, we ensure that content that is in the process of being fact-checked or which cannot be verified is not recommended for users.

TikTok continually invests in enhancing our machine learning and detection capabilities, growing our centralized databases of fact-checked claims and implementing our proactive detection program to improve the identification and removal of misinformation. As of September 2022, the program helped discover 33 new misinformation trends, leading to the takedown of 58,000 videos. We also collaborate closely with safety partners (such as our Content Advisory Council, researchers, civil society groups, and media literacy experts) to help strengthen policies and elevate authoritative voices on our platform.

By way of example, TikTok recently partnered with Melanoma Institute Australia to raise awareness about the dangers of tanning. As part of this campaign, TikTok redirected users who searched for relevant terms such as 'sunburn' or 'tanning' to verified educational material over the length of the campaign.

Objective 4: Empower consumers to make better informed choices of digital content.

Outcome 4: Users are enabled to make more informed choices about the source of news and factual content accessed via digital platforms and are better equipped to identify Misinformation.

Our approach to informing users on identifying misleading information is double-pronged. Firstly, in line with our policies, we do not accept paid political advertisements and educate users on that fact through public messaging. Secondly, we put significant efforts, including in-app features, to provide timely, accurate and authoritative information regarding major civic events, such as elections, public health issues and social concerns. This empowers users to act on the information they see: from inquiring further and seeking authoritative information to reporting suspicious content.

As part of efforts to combat misinformation during the 2022 Victorian State Election, TikTok implemented hashtag PSAs. When users searched or viewed content featuring hashtags related to the elections during the period, a PSA reminded users of our community guidelines and provided contextual information, such as links to the Victoria Electoral Commission's website.

By actively engaging with the Australian Electoral Commission, TikTok launched a dedicated Election Guide in March 2022 to provide trusted and independent information to the Australian community. The Guide:

- promoted the importance of being enrolled to vote;
- provided authoritative information on the election process.

The Election Guide was accessible by searching for election-related content or through a banner on TikTok's Discovery page. It was accessed 107,218 times during election period. In addition, we added PSAs to election-related hashtags, encouraging users to adhere to our Community Guidelines, fact-check new information and report misinformation using a dedicated button within the app. The Australian federal election's PSA received ~90,000 views and is part of our broader efforts to combat misinformation by empowering users to assess information during significant events and take action as appropriate (such as reporting content, researching it or following prompts to authoritative sources of civic information, such as the Australian Electoral Commission).

Objective 5: Improve public awareness of the source of Political Advertising carried on digital platforms.

TikTok

Outcome 5: Users are better informed about the source of Political Advertising.

As an entertainment platform, TikTok takes preventive and proactive measures to ensure integrity of any political content. TikTok strictly prohibits political and issue-based advertising, as per [TikTok advertising policies](#). At the same time, we are committed to enabling authentic conversations on political issues, as well as promoting election integrity through [user education and external partnerships](#).

In 2022, TikTok [extensively supported](#) Australian federal and Victorian state elections through:

- launching the federal Election Guide, along with the Australian Electoral Commission
- connecting users to governmental portals with details on locations, eligibility and voting guidelines
- removing harmful misinformation related to ongoing elections
- adding Public Service Announcements (PSA) to election-related hashtags, encouraging users to report violations and abide by Community Guidelines
- simplifying the user reporting process through a specific election misinformation reporting button.

TikTok also introduced a range of [account-based policies](#) to ensure authenticity and restrict monetisation of [Government, Politician and Political Parties' Accounts](#) (GPPPA):

- Despite not actively encouraging GPPPA to join the platform, we support account [verification](#) to let the users know they are watching authentic content of a governmental authority.
- GPPPA by default do not have access to advertising, monetisation and fundraising features, preventing political accounts from financial gain or user solicitation through TikTok. This includes both political campaign fundraising and use of monetisation tools such as gifts, tips and e-commerce.
- In limited circumstances, GPPPA may require access to ad services to support public health or safety (e.g. promoting COVID-19 boosters). Such campaigns can only be run in close collaboration with TikTok representatives.

Disruption of monetisation incentives for political content and GPPPA verification helps TikTok enable political discussions in a safe, entertaining environment.

Objective 6: Strengthen public understanding of Disinformation and Misinformation through support of strategic research.

Outcome 6: Signatories support the efforts of independent researchers to improve public understanding of Disinformation and Misinformation.

We are currently supporting a strategic research project to better understand perceptions of misinformation in the Australia-specific context. The project aims to understand the extent to which users perceive themselves and others as vulnerable to the effects of misinformation, as well as their confidence and willingness to report it. Additionally, in 2022 TikTok partnered with third-party [IFCN-certified](#) fact-checking organisations [Agence France Presse](#) and [Australian Associated Press](#) to detect and mitigate the spread of misinformation and disinformation in Oceania markets.

Furthermore, a select number of content creators from Australia and New Zealand, were given focused digital media literacy training by the Australian Associated Press. Creators were given tools to confidently source accurate information, comprehend how to critically analyse it, and then encouraged to utilise those tools to produce educational videos which outlined these skills for their own followers, using a "train the trainer" model. The initiative had a broad reach (with creators having a diverse audience of over 3.1 million followers) in educating users on how to recognise false information.

TikTok also offers users access to trustworthy resources through a variety of online tools, in-app PSAs and information centres. These efforts are created in collaboration with respectable third-parties and concentrate on a variety of crucial subjects including COVID-19, elections, mental health and first nations issues.

Objective 7: Signatories publicize the measures they take to combat Disinformation and Misinformation.

Outcome 7: The public can access information about the measures Signatories have taken to combat Disinformation and Misinformation.

TikTok is commited to transparent reporting. To ensure clarity of TikTok's practices for users, law enforcement agencies, governments and general public, we continue to maintain our [Transparency and Accountability Centres](#) (TACs), publish [quarterly transparency reports](#) and provide [newsroom updates](#).

Our TACs in Dublin and Los Angeles host hundreds of guests virtually and opened physically, as COVID-19 measures were eased. Furthermore, we have enhanced the frequency of our reporting and broadened the scope of information included in our [Community Guidelines Enforcement Reports](), such as reporting the number of removed suspected under-age-13 accounts and including information on identification and removal of covert influence operations globally.

We have also piloted a beta-version of our [research API]() to provide anonymised, publicly accessible data on TikTok's content and platform activity. This early version will be tested by members of our Content and Safety Advisory Councils who are also experts on misinformation, violent extremism, hateful behavior and emerging technologies to ensure quick feedback on usability and the overall experience with the system.

Helping our community access authoritative information on the Russia-Ukraine War

As an example of our initiatives to counteract misinformation, we have published a [comprehensive account of our actions during the Russia-Ukraine war]() online.

The war accelerated roll-out of our [state-affiliated media policy]() to label accounts whose editorial output or decision-making process is subject to governmental control or influence. The aim of labelling state-affiliated media is to provide accurate, transparent, and actionable context when people engage with content from media accounts that may present governmental viewpoint. To ensure the efficacy of this policy, we consulted over 60 media experts, political scientists, academics, international organisations and civil representatives across various regions. Additionally, we sought the advice of our [Content and Safety Advisory Council]() and worked with a leading global media monitor to develop an independently-validated methodology to guide our labelling decisions. As of April 2022, we have effectively labelled content from 49 Russian state-controlled media accounts in addition to identifying 16 misinformation narratives that have been swept from the platform.

Our safety team has been [actively removing videos]() related to the Russia-Ukraine war, with a total of 41,191 videos taken down as of April 2022. 87% of these videos were violating our policies against harmful misinformation, and 78% of violations were identified proactively - prior to user reporting. To further combat misinformation on the platform, TikTok also [partnered with fact-checking]() organisations to assess videos globally. Out of 13,738 videos assessed by April 2022, 5,600 videos were labeled with prompts stating that the content could not be verified by fact-checkers.

We also [took action]() against coordinated efforts to sway public opinions and mislead users about account identities. In total, 6 networks and 204 accounts globally have been removed for these activities between February 24 and March 31, 2022. Additionally, we are actively removing fake accounts and related videos: specifically, 321,784 fake accounts in Russia

and 46,298 fake accounts in Ukraine have been removed, resulting in the removal of 343,961 videos over the period. These efforts demonstrate our commitment to maintain a safe and trustworthy platform for our users.

Concluding Remarks

TikTok is a committed signatory to all Objectives and Outcomes, outlined in the Australian Code of Practice on Disinformation and Misinformation. We will continue delivering annual transparency reports on countering misinformation to ensure the trust and safety of our users.

In addition to the initiatives described in this report, we are working towards providing users with even more transparency on moderation decisions, suggested videos and enforcement policies. Since the beginning of 2023, we have been working on a [new account enforcement system](), allowing users see why their videos are [Ineligible for Recommendation]() on the "For You" feed. This promotes awareness of TikTok's Community Guidelines, enables transparent moderation and helps users create safer, more appropriate content. Similarly, a [recently added feature]() allows users to see why a video is "Recommended" to them, providing clarity on the logic of the recommendation system, while [another new platform function]() enables users to "refresh" their For You feed so that, when activated, a user's feed will resemble that of a new user. Further details on moderation, recommendation systems and recommended videos are published in TikTok's [educational resources](), providing users with transparent information on content operations. We also [reworked the account enforcement process]() to improve safeguarding against repeat offenders, [enhanced transparency on account statuses]() for our creators, as well as outlined [content categories ineligible]() for the "For You" feed to increase awareness of recommendation systems. These efforts are aimed at making not only the posted content, but user experience as a whole, more transparent.

While TikTok is primarily an entertainment platform, we strive to enable creativity in a safe environment, as well as support genuine discussions on global affairs, politics and health. We deeply value the trust of our community and are committed to providing a transparent, authentic platform experience globally.

Appendix

Approach to Disinformation and Misinformation

TikTok [takes action](#) against harmful misinformation, affecting individuals, our community or the larger public regardless of the intent. The measures include removal of misleading content and violative accounts, investment into media literacy, direction of users to authoritative sources and external partnerships with experts.

- **Reducing the spread of misinformation**:

Our [Community Guidelines](#) are in place to remove content that could deceive our users about health, current affairs, civic processes, or safety. If the content cannot be verified as authentic, it may become [ineligible for recommendation](#) and may have an added notice to warn other users. Additional feature also [suggests users to reconsider sharing videos](#) that were found to contain unverified messaging.

We combine technology and human moderation to ensure scalability of enforcement measures. Assessment of misinformation often relies on additional contexts, ongoing developments and special training: we provide enhanced training and tools to specialized misinformation moderators, as well as ensure their access to fact-checking partners.

- **Connecting to authoritative information**:

To combat misinformation spread, we actively promote authoritative information sources. The measures include launches of authoritative third-party information centres, public service announcements on hashtags and search terms, and warning labels on unverified content to encourage seeking official information. In addition, we are committed to supporting election integrity: we [cooperate with electoral commissions and civic society associations](#) to ensure authenticity of resources for user education. Such partnerships help authenticate content on in-app election centres, announce election guidelines, prevent violative searches, add label warnings to videos and PSAs to content and hashtags.

- **Developing global fact-checking program**:

Since 2020, we have developed partnerships with 15 fact-checking organisations globally, supporting over 40 languages: all TikTok's fact-checking partners are verified signatories of the [International Fact-Checking Network's](#) [code of principles](#). The following organisations help TikTok assess misinformation: [Agence France-Presse](#), [Animal Político](#), [Australian Associated Press](#), [Code for Africa](#), [dpa Deutsche Presse-Agentur](#), [Estadão Verifica](#), [Facta](#),

Lead Stories, Logically, Newschecker, Newtral, PolitiFact, Reuters, Science Feedback, and Teyit.

Our fact-checkers do not engage in content moderation: instead, their assessments help evaluate moderated content to ensure it follows our Community Guidelines. Partnerships with fact-checkers further help to improve misinformation detection and enforcement through:

- a proactive detection program to identify new and evolving claims across the internet, to then remove violative narratives from TikTok.
- a database of fact-checked claims to assist misinformation moderators to make quick, accurate decisions.

Trust is the cornerstone of safe user experience, and TikTok is committed to making entertainment genuine, accessible and authentic.

Policy on Harmful Misinformation

Misinformation is defined as content that is inaccurate or false. We will remove misinformation that causes significant harm to individuals, our community, or the larger public regardless of intent. Significant harm includes serious physical injury, illness, or death; severe psychological trauma; large-scale property damage, and the undermining of public trust in civic institutions and processes such as governments, elections, and scientific bodies. This does not include simply inaccurate information, myths, or commercial or reputational harm.

Do not post, upload, stream, or share:

- Misinformation that incites hate or prejudice
- Misinformation related to emergencies that induces panic
- Medical misinformation that can cause harm to an individual's physical health
- Content that misleads community members about elections or other civic processes
- Conspiratorial content including content that attacks a specific person or a protected group, includes a violent call to action, or denies a violent or tragic event occurred
- Digital Forgeries (Synthetic Media or Manipulated Media) that mislead users by distorting the truth of events and cause significant harm to the subject of the video, other persons, or society

Do not:

- Engage in coordinated inauthentic behavior such as the use of multiple accounts to exert influence and sway public opinion while misleading individuals, our community,

23

or our systems about the account's identity, location, relationships, popularity, or purpose