Aotearoa New Zealand Code of Practice for Online Safety and Harms 2024 Annual Compliance Report

Signatory:	TikTok
If applicable: Relevant Products / Services:	TikTok
Reporting Period:	1 July 2023 to 30 June 2024

4.1 Reduce the prevalence of harmful content online

Signatories commit to implementing policies, processes, products and/or programs that would promote safety and mitigate risks that may arise from the propagation of harmful content online, as it relates to the themes identified in section 1.4.

Outcome 1. Provide safeguards to reduce the risk of harm arising from online **child sexual exploitation & abuse (CSEA)**

Measure 1. Implement, enforce and/or maintain policies, processes, products, and/or programs that seek to prevent known child sexual abuse material from being made available to users or accessible on their platforms and services. [Opted in]

Measure 2. Implement, enforce and/or maintain policies, processes, products, and/or programs that seek to prevent search results from surfacing child sexual abuse material. [Opted in]

Measure 3. Implement, enforce and/or maintain policies, processes, products, and/or programs that seek to adopt enhanced safety measures to protect children online from peers or adults seeking to engage in harmful sexual activity with children (e.g. online grooming and predatory behaviour). [Opted in]

Measure 4. Implement, enforce and/or maintain policies, processes, products, and/or programs that seek to reduce new and ongoing opportunities for the sexual abuse or exploitation of children. [Opted in]
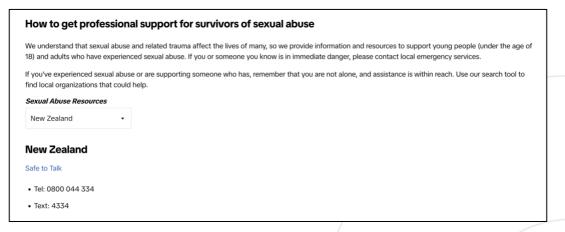
Measure 5. Work to collaborate across industry and with other relevant stakeholders to respond to evolving threats. [Opted in]

TikTok response:

Updates to Policies and Work undertaken during the reporting period:

- We do not allow content that may put young people at risk of psychological, physical, or developmental harm. If we become aware an account holder has a severe violation or has committed a sexual offense against a young person, we will ban the account, as well as any other accounts belonging to that person. This includes content with animation or digitally created or manipulated media, that depicts abuse, exploitation, or the endangerment of minors.

- We report incidents of youth sexual abuse and exploitation to the National Center for Missing and Exploited Children (NCMEC). We also report to relevant law enforcement authorities when there is a specific, credible, and imminent threat to human life or serious physical injury.

- In 2024, we further enhanced our capabilities to protect our users from the harms of Child Sexual Abuse Material (CSAM) through the launch of a search intervention, which alerts users when they are searching for terms or phrases which may be associated with sexualised content of minors. The alert also provides an explicit warning to users reminding them of the severe consequences of creating, viewing, or sharing such content. Users are also directed to self-help resources by Safe to Talk and the New Zealand Department of Internal Affairs.

- Our Safety Centre web landing page for Sexual Abuse Support was similarly updated in early 2024, providing easily navigable contact points for New Zealand users who might require support, with re-direction to Safe to Talk, as highlighted below:

> **How to get professional support for survivors of sexual abuse**
>
> We understand that sexual abuse and related trauma affect the lives of many, so we provide information and resources to support young people (under the age of 18) and adults who have experienced sexual abuse. If you or someone you know is in immediate danger, please contact local emergency services.
>
> If you've experienced sexual abuse or are supporting someone who has, remember that you are not alone, and assistance is within reach. Use our search tool to find local organizations that could help.
>
> *Sexual Abuse Resources*
>
> New Zealand ▾
>
> **New Zealand**
>
> Safe to Talk
>
> - Tel: 0800 044 334
> - Text: 4334

- At TikTok, we recognise the challenge guardians face parenting in the digital age. Keeping teens educated and safe requires collaboration among parents, experts, governments, and platforms. We regularly update the TikTok Guardian's Guide, first launched in May 2021, which is a one-stop-shop to learn about the safety features available to parents and caretakers, as well as their teens. It has information on how to get started on the platform, our safety and privacy tools, and resources to address common online safety questions. It also details our Family Pairing feature which allows parents or caretakers to link their personal account with the content and privacy settings of their teen's TikTok account, with screentime management tools, the ability to mute notifications, and settings to control the content a teen sees. We understand that every teen and family is different, which is why we offer a range of choices and settings to empower parents and caregivers who are looking for easy ways to engage in their teens' life online.

- In 2024, TikTok launched a New Zealand-specific version of the Guardian's Guide with physical copies included in newspapers in New Zealand across 2024.
- Our 'Teen Safety Center' is designed to help younger people understand the different features and safety tools available on TikTok, and empowers them information to help create a positive experience on our platform and with our community.
- In March 2024 we announced the launch of TikTok's global Youth Council, a new initiative that further strengthens how we build our app to be safe for teens by design. The Youth Council was created in partnership with Praesidio Safeguarding, a specialist online safety agency, and is comprised of 15 teens aged between 15 and 18 representing a range of communities and countries from across the globe.
- We conducted in-person discussions and training sessions with the Department of Internal Affairs' Digital Child Exploitation Team and the New Zealand Police's Online Child Exploitation Across New Zealand (OCEANZ) team. During these sessions, we provided guidance on how TikTok can support investigations into Child Sexual Exploitation and Abuse (CSEA) cases and explored opportunities for future collaboration, particularly in relation to insights and signal sharing.
- In 2024, TikTok launched a dedicated page on our Transparency Center to share updates on our efforts to identify and report instances of CSEA.

The table below shows the volume of content removed for violating our Youth Exploitation and Abuse policies in New Zealand for the reporting period, as well as our performance metrics on removal efficiency. Global figures for these and other metrics are published quarterly in TikTok's Community Guidelines Enforcement Reports.

Total Videos Removed	Removal before content is reported by users	Removal within 24 hours of content posted to platform
13,017	97.5%	86.6%

TikTok continues to support and implement the following:

- We are deeply committed to TikTok being a safe and positive experience for all users and particularly people under the age of 18. TikTok is a platform designed with the safety of minors front of mind, and we have a range of policies, processes, and enhanced safety measures in place to protect the safety of minors on TikTok. These include measures to detect, prevent and report the sexual exploitation of minors and grooming behaviour, policies prohibiting content containing nudity and sexual activity involving minors, and minimum age requirements to use TikTok, as stipulated in our Terms of Service.
- We use a combination of AI and human moderation to detect, remove and, where appropriate, report any content that depicts, promotes, normalises, or glorifies grooming behaviours, as well as content that solicits real-world contact between a minor and an adult or between minors with a significant age difference.
- Users can also report content (videos, comments, direct messages, hashes, and audio) if they believe it violates our Community Guidelines. Users can report content in-app and through our website by choosing a reason why they think the content might be violative.

3

- Where we identify that a user has been convicted of crimes against children, we do not allow them to have an account on our platform. These crimes include: sexual assault, molestation, murder, physical abuse or neglect, abduction, international parental kidnapping, trafficking, exploitation of minors for prostitution, live online sexual abuse of a minor, sexual exploitation of minors in the context of travel and tourism, attempts to obtain or distribute CSAM, and the production, possession, or distribution of CSAM.
- Account holders who are under the age of 16 cannot use direct messaging and their content is not eligible to appear in the For You feed.
- Account holders who are under the age of 18 cannot host a livestream and cannot send or receive gifts via our virtual gifting features in livestreams.

Outcome 2: Provide safeguards to reduce the risk of harm arising from online **bullying or harassment**

Measure 6. Implement, enforce and/or maintain policies and processes that seek to reduce the risk to individuals (both minors and adults) or groups from being the target of online bullying or harassment. [Opted in]

Measure 7. Implement and maintain products and/or tools that seek to mitigate the risk of individuals or groups from being the target of online bullying or harassment. [Opted in]

Measure 8. Implement, maintain and raise awareness of product or service related policies and tools for users to report online bullying or harassment content. [Opted in]

Measure 9. Support or maintain programs, initiatives or features that seek to educate and raise awareness on how to reduce or stop online bullying or harassment. [Opted in]

TikTok response:

Updates to Policies and Work undertaken during the reporting period:

- TikTok proactively detects emerging cyberbullying trends through media monitoring, content analysis, and moderator feedback. We implement preventative and mitigative strategies to manage risk and curb harmful trends on the platform and continue to invest heavily in training technology and human moderators to detect, review, and remove harmful content, and continuously optimise our processes to react faster and more effectively. This includes improving our moderation capabilities to better detect and enforce harassment and bullying, and hate speech content.
- TikTok's 'Comment Care Mode' allows comment controls and filters that enable users to restrict who can comment on their content, bulk delete comments, and automatically block specific keywords or "filter all" comments.
- In addition, our 'Rethink' feature provides automated prompts that encourage people to consider the impact of their words before posting a potentially unkind or violative comment.
- We have also launched a dedicated online webpage on bullying prevention on the TikTok Safety Centre. This webpage provides our users with information and resources to help users identify, address and report bullying and harassment.

4

The table below shows the volume of content removed for violating our Harassment and Bullying policies in New Zealand for the reporting period, as well as our performance metrics on removal efficiency.

Total Videos Removed	Removal before content is reported by users	Removal within 24 hours of content posted to platform
27,087	**89.4%**	**88.6%**

TikTok continues to support and implement the following:

- TikTok does not tolerate, and has implemented robust content policies and safeguards against, bullying, shaming, and harassment (including sexual harassment).
- Our Community Guidelines prohibit content and behaviour that expresses abuse, including threats or degrading statements intended to mock, humiliate, embarrass, intimidate, or hurt an individual. This includes content that threatens to hack, 'dox' or blackmail another individual, as well as content that glorifies, normalises, or promotes sexual harassment, regardless of the user's intent. These prohibitions extend to the use of all TikTok features and content.
- To enable good faith expression about matters of public interest, critical comments of public figures may be allowed, however, serious abusive behaviour against public figures is prohibited.
- In terms of age appropriate settings, the 'allow comments on videos' setting for younger teens (13-15) is set to 'Friends' by default and the comment filter for spam and offensive comments is always switched on for this age group, while older teens (16-17) will have this setting on by default.
- Users can also report all content (videos, comments, direct messages, hashes, and sounds) if they believe it violates our Community Guidelines. Users can report content in-app and through our website by choosing a reason why they think the content might be violative.

Outcome 3: Provide safeguards to reduce the risk of harm arising from online **hate speech**

Measure 10. Implement, enforce and/or maintain policies and processes that seek to prohibit or reduce the prevalence of hate speech. [Opted in]

Measure 11. Implement and maintain products and tools that seek to prohibit or reduce the prevalence of hate speech. [Opted in]

Measure 12. Implement, maintain and raise awareness of product or service related policies and tools for users to report potential hate speech. [Opted in]

Measure 13. Support or maintain programs and initiatives that seek to encourage critical thinking and educate users on how to reduce or stop the spread of online hate speech. [Opted in]

Measure 14. Work to collaborate across industry and with other relevant stakeholders to support efforts to respond to evolving harms arising from online hate speech. [Opted in]

TikTok response:

Updates to Policies and Work undertaken during the reporting period:

- We collaborate with government, industry partners, and other relevant stakeholders across New Zealand to proactively prevent and respond to online hate speech. We also expanded our Safety Advisory Council to include representatives from New Zealand, specifically the former Chief Censor, David Shanks, who has been advising TikTok on policy, platform safety, and online harms throughout 2024 including in-person at the APAC Safety Week in Singapore.

- In 2024, we engaged with Te Raranga, a New Zealand Police-led initiative to reduce harm caused by hate in Aotearoa. We shared resources including our Guardian's Guide aimed at Whanau and also our specific safety tools to provide advice on how users can enhance their own protection. We continue to engage with and explore future collaboration opportunities with Te Raranga.

- As in previous years, we recognise the significance of the date of the Christchurch mosque shootings on 15 March. TikTok teams undertook proactive monitoring in the lead up to and on 15 March, to ensure rapid identification and removal of content in violation of our Community Guidelines.

- TikTok rolled out refreshed Community Guidelines in 2024 and began to enforce expanded hate speech and hateful behaviour policies. These policies aim to better address implicit or indirect hate speech and create a safer and more civil environment for everyone. These add to our long-standing policies aimed at tackling hateful ideologies such as antisemitism, Islamophobia, and other forms of hate. For example, we updated our hate speech policy to recognise content that uses "Zionist" as a proxy for a protected attribute when it is not used to refer to a political ideology (and instead for Jewish or Israeli identity). This policy was implemented early this year after observing a rise in how the word was increasingly used in a hateful way.

- In relation to the ongoing war in the Middle East, between 7 October 2023 and 15 September 2024, we removed more than 4.7 million videos globally and suspended more than 300,000 livestreams in the conflict region for violating our Community Guidelines, which included content promoting terrorist groups, hate speech, and misinformation.

- Please also refer to our response in Outcome 2 on the new initiatives that TikTok has undertaken during the reporting period.

The table below shows the volume of content removed for violating our Hate Speech and Hateful Behaviour policies in New Zealand for the reporting period, as well as our performance metrics on removal efficiency.

Total Videos Removed	Removal before content is reported by users	Removal within 24 hours of content posted to platform
11,673	**89.8%**	**79.9%**

TikTok continues to support and implement the following:

- TikTok is a diverse and inclusive community that has no tolerance for discrimination. Our Community Guidelines do not permit content that contains hate speech or involves hateful behaviour, or which praises, promotes, glorifies or supports any hateful ideology (e.g. white supremacy, misogyny, anti-LGBTQ, Islamophobia, and antisemitism). We ban accounts and/or users that engage in severe or multiple hate speech violations or that are associated with hate speech off the TikTok platform.
- As part of updates to our Community Guidelines, TikTok has added 'tribe' as a protected attribute in our hate speech and hateful behaviour policies.
- Please also refer to our response in Outcome 2 on the initiatives that TikTok continues to support and implement.

Outcome 4: Provide safeguards to reduce the risk of harm arising from online **incitement of violence**

Measure 15. Implement, enforce and/or maintain policies and processes that seek to prohibit or reduce the prevalence of content that potentially incites violence. [Opted in]

Measure 16. Implement and maintain products and tools that seek to prohibit or reduce the prevalence of content that potentially incites violence. [Opted in]

Measure 17. Implement, maintain and raise awareness of product or service related policies and tools for users to report content that potentially incites violence. [Opted in]

Measure 18. Support or maintain programs and initiatives that seek to educate users on how to reduce or stop the spread of online content that incites violence. [Opted in]

Measure 19. Work to collaborate across industry and with other relevant stakeholders to support efforts to respond to evolving harms arising from online content that incites violence. [Opted in]

TikTok response:

The table below shows the volume of content removed for violating our Violent and Hateful Organisations and Individuals policies in New Zealand for the reporting period, as well as our performance metrics on removal efficiency.

Total Videos Removed	Removal before content is reported by users	Removal within 24 hours of content posted to platform
10,083	95.8%	83.4%

TikTok continues to support and implement the following:

- TikTok maintains a zero-tolerance policy on violent extremism. Our Community Guidelines clearly outline that we do not allow people to use our platform to threaten or incite violence, or to promote violent extremist organisations, individuals, or acts. When there is a threat to public safety or an account is used to promote or glorify off-platform violence, we ban the account. When warranted, we will report threats to relevant legal authorities.

- To effectively protect our community, we may consider off-platform behaviour to identify violent extremist organisations and individuals on our platform. We do not allow organisations or individuals on our platform who promote or engage in violence, including terrorist organisations, organised hate groups, criminal organisations, and other non-state armed groups that target civilians. If we find such organisations or individuals, we will ban their accounts.

- TikTok takes an uncompromising stance against enabling violent extremism on or off our platform. To further strengthen our commitment to user safety and human rights, we have partnered with Tech Against Terrorism, which brings together technology companies, civil society, and academics over the shared goal of countering violent extremism.

- In addition, our Trust and Safety teams partner with local experts and civil society organisations to understand the unique cultures and experiences of communities affected by violent extremism. We take into account publicly available information from experts, including the United Nations Security Council and Southern Poverty Law Center, to designate dangerous or hateful individuals and organisations.

- Users can also report all content (videos, comments, direct messages, hashes, and sounds) if they believe it violates our Community Guidelines. Users can report content in-app and through the web-app by choosing a reason why they think the content might be violative.

Outcome 5: Provide safeguards to reduce the risk of harm arising from online **violent or graphic content**

Measure 20. Implement, enforce and/or maintain policies and processes that seek to prohibit and/or reduce the spread of violent or graphic content online. [Opted in]

Measure 21. Implement and maintain products and tools that seek to and/or reduce the spread of violent or graphic content. [Opted in]

Measure 22. Implement, maintain and raise awareness of product or service related policies and tools for users to report potential violent and graphic content. [Opted in]

TikTok response:

The below table shows the volume of content removed for violating our Shocking and Graphic Content policies in New Zealand for the reporting period, as well as our performance metrics on removal efficiency.

Total Videos Removed	Removal before content is reported by users	Removal within 24 hours of content posted to platform
20,157	**98.2%**	**95.1%**

TikTok continues to support and implement the following:

- Our Community Guidelines do not allow content that is gratuitously shocking, graphic, sadistic or gruesome, or content that promotes, normalises, or glorifies extreme violence or suffering on our platform. TikTok uses a combination of AI and human moderation to proactively identify and prevent the spread of such content. When it is a threat to public safety, we ban the account and, when warranted, we will report it to relevant legal authorities.
- At the same time, we recognise that some content that would normally be removed pursuant to our Community Guidelines may be in the public interest. Therefore, we may allow exceptions under certain limited circumstances, such as educational, documentary, scientific, artistic, or satirical content, content in fictional or professional settings, counterspeech, or content that otherwise enables individual expression on topics of social importance. To minimise the potentially negative impact of graphic content, we may first include safety measures such as an "opt-in" screen or warning.
- Users can report all content (videos, comments, direct messages, hashes, and sounds) if they believe it violates our Community Guidelines. Users can report content in-app and through the web-app by choosing a reason why they think the content might be violative.

Outcome 6: Provide safeguards to reduce the risk of harm arising from online **misinformation**

Measure 23. Implement, enforce and/or maintain policies, processes and/or products that seek to reduce the spread of online misinformation. [Opted in]

Measure 24. Implement, enforce and/or maintain policies and processes that seek to penalise users who repeatedly post or share misinformation that violates related policies. [Opted in]

Measure 25. Support or maintain media literacy programs and initiatives that seek to encourage critical thinking and educate users on how to reduce or stop the spread of misinformation. [Opted in]

Measure 26. Support or maintain programs and/or initiatives that seek to support civil society, fact-checking bodies and/or other relevant organisations working to combat misinformation. [Opted in]

Measure 27. Work to collaborate across industry and with other relevant stakeholders to support efforts to respond to evolving harms arising from misinformation. [Opted in]

TikTok response:

Updates to Policies and Work undertaken during the reporting period:

- TikTok's most recent update to our Community Guidelines in April 2024 added and updated definitions to provide more clarity about our policies, for example on health misinformation.

- As more creators take advantage of AI to enhance their creativity, we want to support transparent and responsible content creation practices. As of September 2023, we launched a new a "AI-generated" label for creators to disclose content that is completely AI-generated or significantly edited by AI. Additionally, AI-generated content (AIGC) made with TikTok's in-house AI effects is automatically labelled. To date, more than 37 million creators have used TikTok's AIGC labelling tool.

- Additionally, as AI evolves, TikTok continues advancing detection through new technologies. In May 2024, we introduced auto-labels for AIGC that has Coalition for Content Provenance and Authenticity (C2PA) Content Credentials attached. TikTok is the first platform to implement C2PA Content Credentials, a technology developed by Coalition for Content Provenance and Authenticity. It tags content with metadata that tells us it's AI-generated.

- We work to protect the integrity of elections by removing or reducing the reach of harmful misinformation, which we define as false or misleading content, regardless of intent. This includes false or misleading information about how to vote, how to register to vote, the eligibility qualifications for candidates, and the procedures that govern the implementation of elections. We detect and take action against such content by using a combination of AI and human moderators, and by working with Australian Associated Press (AAP) and other International Fact-Checking Network-certified fact-checking experts to assess the accuracy of content.

- For the New Zealand General Election on 14 October 2023, TikTok partnered with Elections NZ to launch proactive front-end informative features such as in-app Public Service Announcements or PSAs (search guide interventions) that will prompt users to refer to official resources when engaging with election-related information. This is in addition to various risk mitigation measures we implemented to protect the integrity of the 2023 New Zealand General Election, which included removing accounts impersonating politicians, political parties, and government organisations. We also worked with an external specialised language-support partner to support us, as required, with the review of Māori-related content on TikTok.

The below table shows the volume of content removed for violating our Misinformation policies in New Zealand for the reporting period, as well as our performance metrics on removal efficiency.

Total Videos Removed	Removal before content is reported by users	Removal within 24 hours of content posted to platform
4,241	**95.9%**	**49.9%**

TikTok continues to support and implement the following:

- We do not allow users to post content containing misinformation that causes significant harm to individuals, our community, or the larger public regardless of intent. This includes inaccurate or false content that may cause 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.
- TikTok's Community Guidelines also prohibit the following:
  - Content containing 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
  - 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
- In addition to removing content that is inaccurate and harms our users or community, 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. These actors never stop evolving their tactics, and we continually seek to strengthen our policies as we detect new types of content and behaviours.
- We empower users to make informed decisions about the content and information they consume on TikTok, providing publicly available account information (number of accounts following, followers, and likes) on TikTok profiles and supporting blue-tick verification for public figures and organisations, to help users identify authentic sources of information.
- While TikTok uses a combination of technology and thousands of safety professionals to enforce our Community Guidelines, we also rely on specialised misinformation moderators who have enhanced training, expertise, and tools to take action on misinformation.
- Content which may be potentially misleading is flagged by our moderators and may be removed from TikTok's For You page, while the information is sent to third party fact-checkers for review. Subsequently, content which is deemed false and misleading will be removed in accordance with our Community Guidelines.
- TikTok partners with AAP to detect and mitigate the spread of misinformation and disinformation in Oceania markets.
- TikTok engaged AAP to create and deliver targeted digital media literacy education to a small group of content creators, including from New Zealand. The aim was to empower the participants to confidently source reliable and factual information, understand how to interrogate that information effectively, and encourage them to use those skills to create factual, reliable, and educative content. The creators selected have large followings, with the project aiming to have a wide reach and impact in educating audiences on how to identify mis- and disinformation.
- Users can also report all content (videos, comments, direct messages, hashes, and sounds) if they believe it violates our Community Guidelines. Users can report content in-app and through our website by choosing a reason why they think the content might be violative.

- TikTok also has a range of online resources, in-app Public Service Announcements and hubs (in-app information centres) developed in partnership with reputable third parties focusing on a range of important topics including elections, harmful misinformation, mental health, and other issues that provide users access to reliable information.

Outcome 7: Provide safeguards to reduce the risk of harm arising from online **disinformation**

Measure 28. Implement, enforce and/or maintain policies, processes and/or products that seek to suspend, remove, disable, or penalise the use of fake accounts that are misleading, deceptive and/or may cause harm. [Opted in]

Measure 29. Implement, enforce and/or maintain policies, processes and/or products that seek to remove accounts, (including profiles, pages, handles, channels, etc.) that repeatedly spread disinformation. [Opted in]

Measure 30. Implement, enforce and/or maintain policies, processes and/or products that seek to provide information on public accounts (including profiles, pages, handles, channels, etc.) that empower users to make informed decisions (e.g. date a public profile was created, date of changes to primary account information, number of followers). [Opted in]

Measure 31. Implement, enforce and/or maintain policies, processes and/or products that seek to provide transparency on paid political content (e.g. advertising or sponsored content) and give users more context and information (e.g. paid political or electoral ad labels or who paid for the ad).

Measure 32. Implement, enforce and/or maintain policies, processes and/or products that seek to disrupt advertising and/or reduce economic incentives for users who profit from disinformation. [Opted in]

Measure 33. Work to collaborate across industry and with other relevant stakeholders to support efforts to respond to evolving harms arising from disinformation. [Opted in]

TikTok response:

Updates to Policies and Work undertaken during the reporting period:

- We continuously work to detect and disrupt covert influence operations that attempt to establish themselves on TikTok and undermine the integrity of our platform. Starting in January 2024, we have published detailed information each month on our efforts to disrupt covert influence operations on TikTok on our Transparency Centre webpage, in order to promote transparency and share learnings with our industry.
- Please also refer to our response in Outcome 6 on the new initiatives that TikTok has undertaken during the reporting period.

TikTok continues to support and implement the following:

- We do not allow content or activities that facilitate the spread of disinformation, or which may otherwise undermine the integrity of our platform or the authenticity of our users. In addition to our Community Guidelines prohibiting harmful misinformation, we also remove content or accounts that involve spam, fake engagement, impersonation, and coordinated inauthentic behaviour, 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.
- As part of our Community Guidelines, we have updated our rules regarding how we treat synthetic media, which is content created or modified by AI technology. Synthetic or manipulated media that shows realistic scenes must be clearly disclosed and synthetic media that contains the likeness of any real private figure is not allowed. This update to our Community Guidelines also provided more detail about our work to protect civic and election integrity, including our approach to government, politician and political party accounts.
- Our strict advertising policies exceed industry standards, and all ads must undergo a review process which involves vetting the products/services promoted, ad caption, text, images, audio, visuals, age/region targeting, and landing pages. We do not allow the promotion, sale, or solicitation or facilitation of access to products or services that might be or are considered deceptive, misleading, or unlawful, including unwarranted claims, misinformation including pricing/discount or promotion information inconsistency, and missing T&Cs or privacy policy pages. We also require 18+ targeting for certain categories of products and service and proper disclaimers must be included when applicable.
- Please also refer to our response in Outcome 6 on the initiatives that TikTok continues to support and implement.

Regarding **Measure 31:**

- Paid political advertising, including sponsored political content, is strictly prohibited on TikTok. TikTok does not allow anyone to place political ads, nor do we allow politicians and political party accounts to place ads. We also prevent governments, politicians and political party accounts from accessing our monetisation features and campaign fundraising.
- Sharing political beliefs and engaging in political conversation is allowed as organic content, but our policies prohibit users from paying to advertise or promote this content. We allow some cause-based advertising and public service advertising from government agencies, non-profits, and other entities, provided they are not politically partisan and make exceptions for governments in certain circumstances, e.g., to promote public health messaging.

4.2 Empower users to have more control and make informed choices

Signatories recognise that users have different needs, tolerances, and sensitivities that inform their experiences and interactions online. Content or behavior that may be appropriate for some will not be appropriate for others, and a single baseline may not adequately satisfy or protect all users. Signatories will therefore empower users to have control and to make informed choices over the content they see and/or their experiences and interactions online. Signatories will also provide tools, programs, resources and/or services that will help users stay safe online.

Outcome 8. Users are empowered to **make informed decisions** about the content they see on the platform

Measure 34. Implement, enforce and/or maintain policies, processes, products and/or programs that helps users make more informed decisions on the content they see. [Opted in]

Measure 35. Implement, enforce and/or maintain policies, processes, products and/or programs that seek to promote accurate and credible information about highly significant issues of societal importance and of relevance to the digital platform's user community (e.g. public health, climate change, elections). [Opted in]

Measure 36. Launch programs and/or initiatives that educate or raise awareness on disinformation, misinformation and other harms, such as via media/digital literacy campaigns. [Opted in]

TikTok response:

Updates to Policies and Work undertaken during the reporting period:

- Please also refer to our response in Outcome 6 on the new initiatives that TikTok has undertaken during the reporting period.

TikTok continues to support and implement the following:

- TikTok present users with a stream of videos on our For You feed curated to their interests, making it easy to find content and creators they love. The For You feed is powered by a recommendation system that delivers content to each user that is likely to be of interest to that particular user, but also works to intersperse recommendations that might fall outside people's expressed preferences, offering an opportunity to discover new categories of content. We also provide a tool to help people understand why a particular video has been recommended to them.
- The For You feed can also be curated by users. If a video is not quite to a user's taste, we empower our users to long-press on a video and tap "Not Interested" to indicate that they don't care for a particular video. We have also finished rolling out a feature that lets people choose words or hashtags associated with content they don't want to see in their For You feed, to offer another way to help people customise their feed.
- Our recommendation system on TikTok is also designed with safety as a key consideration. Reviewed content found to depict things like graphic medical procedures or the legal consumption of regulated goods, for example – which may be shocking if surfaced as a recommended video to a general audience that hasn't opted in to such content – may not be eligible for recommendation.

TikTok

Outcome 9. Users are **empowered with control** over the content they see and/or their experiences and interactions online

Measure 37. Implement, enforce and/or maintain policies, processes, products and/or programs that seek to provide users with appropriate control over the content they see, the character of their feed and/or their community online. [Opted in]

Measure 38. Launch and maintain products that provide users with controls over the appropriateness of the ads they see.

TikTok response:

Updates to Policies and Work undertaken during the reporting period:

- Please refer to our response in Outcome 1 on the new initiatives that TikTok has undertaken during the reporting period.

TikTok continues to support and implement the following:

- TikTok provides users with tools that they can use to automatically filter out videos with words or hashtags they don't want to see in their For You or Following feeds e.g. a vegetarian who wants to see fewer meat recipes.
- Users can also use a 'Refresh Your Feed' feature that allows them to reset their For You feed if their recommendations no longer feel relevant. When enabled, this feature allows someone to view content on their For You feed as if they just signed up for TikTok. Our recommendation system will then begin to surface more content based on new interactions.
- We have comment controls and filters that enable users to restrict who can comment on their content, bulk delete comments and automatically block specific keywords or "filter all" comments. In addition, our Rethink feature provides automated prompts that encourage people to consider the impact of their words before posting a potentially unkind or violative comment.
- We also empower TikTok users with other various privacy and safety features to control their interactions on the app by:
    - Blocking the accounts they don't want to interact with
    - Setting their account private, where only approved users can follow them and watch their content
    - Choosing if they wish to receive Direct Messages (DMs) or not
    - Choosing who can comment on their content or even turning off comments on their content altogether
    - Filtering comments containing any keywords they don't approve of
    - Hiding all the comments on their content until they review and approve them
    - Choosing who can tag them in their content or even prohibiting anyone from tagging them in their content
    - Choosing who can mention them in their content or even prohibiting anyone from mentioning them in their content
    - Choosing if anyone can duet with their videos
    - Choosing who can see their following list i.e. people they follow

Regarding **Measure 38**:

TikTok continues to support and implement the following:

- TikTok enforces strict advertising policies to protect all users on our platform. We place considerable emphasis on proactive content moderation and the vast majority of the violative content we remove is taken down before it is reported to us or receives any views.
- All ads must undergo a review process which involves vetting the products/services promoted, ad caption, text, images, audio, visuals, age/region targeting, and landing pages.
- We have an extensive list of prohibited products and services that cannot be advertised on the platform including: tobacco, drugs, adult services, weight loss management/supplements, and political advertising.
- We take a special level of care and caution when it comes to advertising that our younger users see and craft our policies to ensure that any ads that could be shown to younger audiences are safe for those viewers.
- The broadest category of protections for minors in TikTok ads starts with the platform's outright ban of any advertising that appeals directly to children, either by influencing children directly/indirectly or appealing to children to get their parents to buy a product. Ads for children's toys and clothing are allowed provided they are targeted and appeal to adults.
- Beyond this broad protection, we moderate ads for any products or services that could potentially pose a higher physical, emotional, or financial risk to minors and restrict them to an 18+ audience. Examples include dating apps, restricting financial services and health products.
- Additionally, we understand that our platform's products encourage mimicry, so we've taken minor safety into consideration and limit the behaviours that can be shown in ads well beyond what is legally required (banning ads showing unsafe driving, dangerous stunts, etc.).
- We regularly review our advertising policies, and our Community Guidelines, to ensure they are keeping up with the development of new products and services and cultural trends, and keep our users safe on TikTok.

TikTok

4.3 Enhance transparency of policies, processes and systems

Transparency helps build trust and facilitates accountability. Signatories will provide transparency of their policies, processes and systems for online safety and content moderation and their effectiveness to mitigate risks to users. Signatories, however, recognise that there is a need to balance public transparency of measures taken under the Code with risks that may outweigh the benefit of transparency, such as protecting people's privacy, protecting trade secrets and not providing threat actors with information that may expose how they may circumvent or bypass enforcement protocols or systems.

Outcome 10. Transparency of policies, systems, processes and programs that aim to reduce the risk of online harms

Measure 39. Publish and make accessible for users Signatories' safety and harms-related policies and terms of service. [Opted in]

Measure 40. Publish and make accessible information (such as via blog posts, press releases and/or media articles) on relevant policies, processes, and products that aim to reduce the spread and prevalence of harmful content online. [Opted in]

TikTok response:

Updates to Policies and Work undertaken during the reporting period:

- Please refer to our response in Outcomes 6 and 7 on the new initiatives that TikTok has undertaken during the reporting period.

TikTok continues to support and implement the following:

- TikTok's Community Guidelines, safety policies and terms of service are publicly available on the TikTok website. These policies cover a broad range of issue verticals including Youth Safety and Well-Being, Safety and Civility, Mental and Behavioral Health, Sensitive and Mature Themes, Integrity and Authenticity and Regulated Goods and Commercial Activities.
- In consultation with relevant stakeholders, we update our Community Guidelines from time to time to evolve alongside new behaviours and risks, as part of our commitment to keeping TikTok a safe place for creativity and joy.
- TikTok's Newsroom, a publicly accessible webpage, outlines information including media articles and other relevant publications which highlight the work being done by the platform to address the spread of harmful online content. This includes updates on our work to counter misinformation on the platform, and our efforts to prevent the spread of violent extremism and its associated ideologies.
- We also publish quarterly Community Guideline Enforcement Reports, with additional Transparency Reports every six months on Government and Law Enforcement requests.

![TikTok logo]

Outcome 11. Publication of regular **transparency reports** on efforts to reduce the spread and prevalence of harmful content and related KPIs/metrics

Measure 41. Publish periodic transparency reports with KPIs/metrics showing actions taken based on policies, processes and products to reduce the spread or prevalence of harmful content (e.g. periodic transparency reports on removal of policy-violating content). [Opted in]

Measure 42. Submit to the Administrator an annual compliance report, as required in section 5.4, that set out the measures in place and progress made in relation to Signatories' commitments under the Code. [Opted in]

TikTok response:

Updates to Policies and Work undertaken during the reporting period:

- Starting in January 2024, we have provide detailed information each month on our efforts to disrupt covert influence operations on TikTok on our Transparency Centre webpage.

TikTok continues to support and implement the following:

- TikTok uses a combination of AI and human moderators to identify, review, and action content that violates our policies. We compile metrics on these actions through our Community Guideline Enforcement reports, which are publicly accessible on the TikTok Transparency Centre webpage, and provide quarterly insights into the volume and nature of content and accounts removed from our platform. We also publish reports on Government and Law Enforcement requests every six months.
- As founding signatories to the Aotearoa New Zealand Code of Practice for Online Safety and Harms, TikTok commits to submitting to the Administrator annual compliance reports that will be used to evaluate our compliance and progress made against relevant outcomes and measures in the Code. TikTok submitted its baseline report in November 2022, annual report in October 2023 and has provided our subsequent annual compliance report for 2024 accordingly.

TikTok

4.4 Support independent research and evaluation

Independent local, regional or global research by academics and other experts to understand the impact of safety interventions and harmful content on society, as well as research on new content moderation and other technologies that may enhance safety and reduce harmful content online, are important for continuous improvement of safeguarding the digital ecosystem. Signatories will seek to support or participate in these research efforts.

Signatories may also seek to support independent evaluation of the systems, policies and processes they have implemented under the commitments of the Code. This may include broader initiatives undertaken at the regional or global level, such as independent evaluations of Signatories' systems.

Outcome 12. Independent research to understand the impact of safety interventions and harmful content on society and/or research on new technologies to enhance safety or reduce harmful content online.

Measure 43. Support or participate, where appropriate, in programs and initiatives undertaken by researchers, civil society and other relevant organisations (such as fact-checking bodies). This may include broader regional or global research initiatives undertaken by the Signatory which may also benefit Aotearoa New Zealand. [Opted in]

Measure 44. Support or convene at least one event per year to foster multi-stakeholder dialogue, particularly with the research community, regarding one of the key themes of online safety and harmful content, as outlined in section 4. This may include broader regional or global events undertaken by the Signatory which involve Aotearoa New Zealand. [Opted in]

TikTok response:

Updates to Policies and Work undertaken during the reporting period:

- We continue to work with a variety of New Zealand-based organisations through our Community Partner Channel (CPC), which is a direct avenue for trusted flaggers to report content to TikTok for review and sits alongside our in-app reporting function. From December 2023, TikTok increased efforts to onboard new partners to the existing New Zealand cohort, inviting organisations, training them in the CPC system, and undertaking a series of in-person sessions with partners in New Zealand across a range of policy areas.
- Please also refer to our response in Outcome 3 on the new initiatives that TikTok has undertaken during the reporting period.

TikTok continues to support and implement the following:

- We are continually assessing ways that we can create safer experiences and better resources for users. The decisions we make in this regard are informed by external research and engagement with experts. Examples of research, partnerships, and multi-stakeholder programs we have supported are set out below.

19

- In October 2023, TikTok co-sponsored the 'Breaking Glass Ceilings and Building Bridges: a trans-Tasman conference exploring the human experience in online spaces and places' event organised by Netsafe New Zealand in association with Australia's eSafety Commissioner. The conference brought together leading online safety experts and practitioners from Aotearoa, Australia, and internationally, to share knowledge and insights, and to participate in discussion and debate, exploring themes focusing on online harm and identity, including:
  - Legislative and policy responses
  - Diversity and inclusion
  - Innovative education
  - Pacific collaboration

Outcome 13. Support independent evaluation of the systems, policies and processes that have been implemented in relation to the Code.

Measure 45. Commit to selecting an independent third-party organisation to review the annual compliance reports submitted by Signatories, and evaluate the level of progress made against the Commitments, Outcomes and Measures, as outlined in section 4, as well as commitments made by Signatories in their Participation Form (see Appendix 2). [Opted in]

TikTok response:

- As a founding signatory to the Aotearoa New Zealand Code of Practice for Online Safety and Harms, TikTok supports the Code's use of a third-party organisation to assess all signatories' (including TikTok's) annual compliance report to evaluate compliance and progress made against outcomes and measures relevant to each platform. TikTok submitted its baseline report in November 2022, our annual report in October 2023, and has provided our subsequent annual compliance report for 2024 accordingly.

20