TikTok

Aotearoa New Zealand Code of Practice for Online Safety and Harms – 2023 Annual Compliance Report

Signatory:	TikTok New Zealand

If applicable: **Relevant Products / Services:**	TikTok

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]

- As explained in our baseline report, our platform is 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.

- Our new Community Guidelines were put into effect April 2023, overhauling how we organise our rules thematically into different topic areas and offering our community more transparency about our rules and how we enforce them. They continue to strictly prohibit activities that enable or perpetuate the abuse, harm, endangerment, or exploitation of minors on TikTok.

- TikTok will take action on any content or accounts involving child sexual abuse material (CSAM) or sexual exploitation of a minor. Any content, including animation or digitally created or manipulated media, that depicts abuse, exploitation, or endangerment of minors is a violation of our Community Guidelines and will be removed when detected. We report CSAM and supporting evidence to the National Center for Missing & Exploited Children (**NCMEC**) and to any additional relevant legal authorities.

- 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 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 violate our Community Guidelines.

- Account holders who are under the age of 16 cannot use direct messaging and need to be at least 18 years old to host a livestream and for their content to be eligible to appear in the For You feed. Account holders who are under the age of 18 cannot send or receive gifts via our virtual gifting features.

- Our Family Pairing feature also allows parents and guardians more control over their younger ones experience on TikTok, allowing them to adjust privacy and experience settings and implement features like our Screen Time Management tool and Private Mode.

- In July 2022, we partnered with Netsafe and Professor Amanda Third to host a webinar to provide an educational session for parents in NZ - this included a strong focus on protecting children online, including education about our family pairing feature, age restriction information and community controls. Amanda is an expert in user-centred, participatory research, and her work investigates children's and young people's technology practices.

Relevant *Global* Community Guidelines Enforcement Statistics:

- In Q1 2023, less than **1%** of all videos uploaded globally were removed for violating our Community Guidelines.

- In Q1 2023, **30.6%** of total videos removed violated our minor safety policies.

    - **98.9%** were proactively removed before any video reports.

    - **88.0%** of these videos were removed before any video views.

    - **91.2%** were removed within 24 hours of being posted.

Relevant *New Zealand* Community Guidelines Enforcement Statistics:

- In Q4 2022, nearly 160,000 videos were removed in New Zealand for violating our Community Guidelines.

    - **95.8%** were proactively removed before any video reports.

    - **78.0%** of these videos were removed before any video views.

    - **88.3%** were removed within 24 hours of being posted.

- In Q1 2023, nearly 120,000 videos were removed in New Zealand for violating our Community Guidelines.

    - **95.1%** were proactively removed before any video reports.

    - **66.1%** of these videos were removed before any video views.

    - **84.5%** were removed within 24 hours of being posted.

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 does not tolerate, and has implemented robust content policies and safeguards against, bullying 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.

- 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.

- 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 our website by choosing a reason why they think the content might violate our Community Guidelines.

- We also 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.

- 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 younger teens (13-15), while older teens (16-17) will have this setting on by default.

- TikTok's Youth Portal and Safety Centre also contains information and resources to help users identify, address and report bullying and harassment.

- In November 2022, TikTok partnered with NGO Bully Zero to run an educative workshop at TikTok's Creator Safety Summit. The workshop focussed on strategies to prevent and reduce the prevalence of online harassment and bullying and included some of TikTok's top creators, including from New Zealand. In the weeks following the workshop, the creators made videos highlighting their key messages and learnings to their followers on the app, including in New Zealand.

- During Netsafety Week 2023, TikTok presented to an online audience of Aotearoa New Zealand TikTok users and other interested parties, ranging from parents to teachers, to raise awareness of features to report content, including content that might be forms of online bullying or harassment.

Relevant *Global* Community Guidelines Enforcement Statistics:

- In Q1 2023, less than **1%** of all videos uploaded globally were removed for violating our Community Guidelines.

- In Q1 2023, **5.3%** of total videos removed violated our harassment and bullying policies.

    - **85.3%** were proactively removed before any video reports.

    - **58.9%** of these videos were removed before any video views.

    - **79.5%** were removed within 24 hours of being posted.

Relevant *New Zealand* Community Guidelines Enforcement Statistics:

- In Q4 2022, nearly 160,000 videos were removed in New Zealand for violating our Community Guidelines.

    - **95.8%** were proactively removed before any video reports.

    - **78.0%** of these videos were removed before any video views.

    - **88.3%** were removed within 24 hours of being posted.

- In Q1 2023, nearly 120,000 videos were removed in New Zealand for violating our Community Guidelines.

    - **95.1%** were proactively removed before any video reports.

    - **66.1%** of these videos were removed before any video views.

    - **84.5%** were removed within 24 hours of being posted.

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]

- Our Community Guidelines, which we actively enforce, do not allow any hateful behaviour, hate speech, or promotion of hateful ideologies. This includes content that attacks a person or group on the basis of their protected attributes.

- As part of updates to our Community Guidelines that came into effect April 2023, TikTok has added 'tribe' as a protected attribute in our hate speech and hateful behaviour policies.

- 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.

- 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 violate our Community Guidelines.

- We collaborate with government, industry partners and other relevant stakeholders across New Zealand to proactively prevent and respond to potential online hate speech. In 2023, this work included proactive moderation efforts in the lead up to major public events such as *Purapura Whetu/Stars of Matariki*, a LIVE event on TikTok and TVNZ Digital+ to celebrate Matariki.

- During Netsafety Week 2023, TikTok presented to an online audience of Aotearoa New Zealand TikTok users and other interested parties, ranging from parents to teachers, to raise awareness of features to report content, including content that they believe might violate our Community Guidelines.

- In November 2022, TikTok partnered with Multicultural NSW (who created the "remove hate from the debate" web tool) to run an educative workshop at TikTok's Creator Safety Summit. The workshop focussed on strategies to prevent and reduce the prevalence of online hate speech and hateful behaviour and included some of TikTok's top creators, including from New Zealand. In the weeks following the workshop, the creators made videos highlighting their key messages and learnings to their followers on the app, including in New Zealand.

Relevant *Global* Community Guidelines Enforcement Statistics:

- In Q1 2023, less than **1%** of all videos uploaded globally were removed for violating our Community Guidelines.

- In Q1 2023, **2.5%** of total videos removed violated our hateful behaviour policies.
    - **89.2%** were proactively removed before any video reports.
    - **74.9%** of these videos were removed before any video views.
    - **83.8%** were removed within 24 hours of being posted.

![TikTok]

Relevant *New Zealand* Community Guidelines Enforcement Statistics:

- In Q4 2022, nearly 160,000 videos were removed in New Zealand for violating our Community Guidelines.

    - **95.8%** were proactively removed before any video reports.

    - **78.0%** of these videos were removed before any video views.

    - **88.3%** were removed within 24 hours of being posted.

- In Q1 2023, nearly 120,000 videos were removed in New Zealand for violating our Community Guidelines.

    - **95.1%** were proactively removed before any video reports.

    - **66.1%** of these videos were removed before any video views.

    - **84.5%** were removed within 24 hours of being posted.

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]

- Our Community Guidelines, which we enforce, make it clear that we do not allow any violent threats, incitement to violence, or promotion of criminal activities that may harm people, animals, or property. If there is a specific, credible, and imminent threat to human life or serious physical injury, we report it to relevant law enforcement authorities.

- We engage regionally and internationally on this important topic, including through multi-party dialogues. In late 2022, TikTok actively participated in the APAC Regional Aqaba Dialogue, alongside other tech companies, international organisations and state representatives (including the New Zealand Government), resulting in a joint communique.

- In September 2022, we became members of Tech Against Terrorism, which brings together technology companies, civil society, and academics over the shared goal of countering violent extremism.

- We also 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, as well as members of the public, can report content in-app or through the web-based form, by choosing a reason why they think the content might violate our Community Guidelines.

- In 2023, TikTok took action to reduce the prevalence of content associated with violent criminal organisations on-platform, such as outlaw motorcycle gangs in Australia and New Zealand, as reported in Australian and New Zealand media.

Relevant *Global* Community Guidelines Enforcement Statistics:

- In Q1 2023, less than **1%** of all videos uploaded globally were removed for violating our Community Guidelines.

- In Q1 2023, **1.4%** of total videos removed violated our violent extremism policies.
    - **94.9%** were proactively removed before any video reports.
    - **77.4%** of these videos were removed before any video views.
    - **85.9%** were removed within 24 hours of being posted.

Relevant *New Zealand* Community Guidelines Enforcement Statistics:

- In Q4 2022, nearly 160,000 videos were removed in New Zealand for violating our Community Guidelines.
    - **95.8%** were proactively removed before any video reports.
    - **78.0%** of these videos were removed before any video views.
    - **88.3%** were removed within 24 hours of being posted.

- In Q1 2023, nearly 120,000 videos were removed in New Zealand for violating our Community Guidelines.
    - **95.1%** were proactively removed before any video reports.
    - **66.1%** of these videos were removed before any video views.
    - **84.5%** were removed within 24 hours of being posted.

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]

- Our Community Guidelines recognise that violent or graphic content may be triggering, cause psychological harm, or lead to extreme discomfort. We do not allow gory, gruesome, disturbing, or extremely violent content. Content is age-restricted if it shows human or animal blood. Content is ineligible for the FYF if it shows fictional violence, blood, potentially distressing or mildly graphic material, animal genitalia or sexual activity between animals.

- TikTok uses a combination of AI and human moderation to proactively identify and prevent the spread of such content on our platform. When it is a threat to public safety, we ban the account and, when warranted, we will report it to relevant legal authorities.

- To help users manage their TikTok experience, we apply an "opt-in" screen or warning information to some content, such as human or animal blood, wild animals attacking each other, professional fighting, content allowed under a public interest exception, and potentially distressing or mildly graphic material.

- 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 violate our Community Guidelines.

- During Netsafety Week 2023, TikTok presented to an online audience of Aotearoa New Zealand TikTok users and other interested parties, ranging from parents to teachers, to raise awareness of features to report content, including content that might be violent and graphic.

Relevant *Global* Community Guidelines Enforcement Statistics:

- In Q1 2023, less than **1%** of all videos uploaded globally were removed for violating our Community Guidelines.

- In Q1 2023, **9.1%** of total videos removed violated our violent and graphic content policies.
  - **97.6%** were proactively removed before any video reports.
  - **82.6%** of these videos were removed before any video views.
  - **92.0%** were removed within 24 hours of being posted.

Relevant *New Zealand* Community Guidelines Enforcement Statistics:

- In Q4 2022, nearly 160,000 videos were removed in New Zealand for violating our Community Guidelines.
  - **95.8%** were proactively removed before any video reports.
  - **78.0%** of these videos were removed before any video views.
  - **88.3%** were removed within 24 hours of being posted.
- In Q1 2023, nearly 120,000 videos were removed in New Zealand for violating our Community Guidelines.
  - **95.1%** were proactively removed before any video reports.
  - **66.1%** of these videos were removed before any video views.
  - **84.5%** were removed within 24 hours of being posted.

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]

- As explained in our baseline report, TikTok does 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.

- Content which may be potentially misleading is flagged by our moderators and may be removed from TikTok's "For You page" (FYP) 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 to our Community Guidelines.

- TikTok partners with third party fact checking organisations to detect and mitigate the spread of misinformation and disinformation in Oceania markets; leveraging the insights of IFCN-certified fact-checking experts.

- We also empower users to make informed decisions about the content and information they consume on TikTok, providing publicly available account information (no. 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.

- In January 2023, we expanded our state-controlled media label program to over 40 markets across multiple regions - ensuring our users have accurate, transparent, and actionable context when they engage with content from media accounts that may present the viewpoint of a government.

- 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 our website by choosing a reason why they think the content might violate our Community Guidelines.

- TikTok also has a range of online resources, in-app PSAs and hubs developed in partnership with reputable third-parties focusing on a range of important topics including COVID-19, elections, and mental health that provide users access to reliable information. In preparation for the General Election on 14 October 2023, TikTok has partnered with Elections NZ to launch in-app PSAs (search guide interventions) that will prompt users to refer to official resources when engaging with election related information.

- In October 2022, TikTok engaged the Australian Associated Press to create and deliver targeted digital media literacy education to a small group of content creators, including from New Zealand. The workshops empowered 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 had large followings, with the project focussed on having a wide impact in educating audiences on how to identify mis- and disinformation.

- In Q3 2023, TikTok is also convening an internal academic/expert roundtable discussing potential online safety challenges, including online misinformation, and recommendations.

Relevant *Global* Community Guidelines Enforcement Statistics:

- In Q1 2023, less than **1%** of all videos uploaded globally were removed for violating our Community Guidelines.

- In Q1 2023, **1.0%** of total videos removed violated our integrity and authenticity policies.

    o **94.8%** were proactively removed before any video reports.

    o **72.8%** of these videos were removed before any video views.

    o **76.6%** were removed within 24 hours of being posted.

- In Q1 2023, TikTok removed more than 51 million fake accounts, 531 million fake likes, 1.2 trillion fake followers globally.

Relevant *New Zealand* Community Guidelines Enforcement Statistics:

- In Q4 2022, nearly 160,000 videos were removed in New Zealand for violating our Community Guidelines.

    o **95.8%** were proactively removed before any video reports.

    o **78.0%** of these videos were removed before any video views.

    o **88.3%** were removed within 24 hours of being posted.

- In Q1 2023, nearly 120,000 videos were removed in New Zealand for violating our Community Guidelines.

    o **95.1%** were proactively removed before any video reports.

    o **66.1%** of these videos were removed before any video views.

    o **84.5%** were removed within 24 hours of being posted.

Relevant *New Zealand* Community Guidelines Enforcement Statistics:

- In Q4 2022, nearly 160,000 videos were removed in New Zealand for violating our Community Guidelines.

    o **95.8%** were proactively removed before any video reports.

    o **78.0%** of these videos were removed before any video views.

    o **88.3%** were removed within 24 hours of being posted.

- In Q1 2023, nearly 120,000 videos were removed in New Zealand for violating our Community Guidelines.

    - **95.1%** were proactively removed before any video reports.

    - **66.1%** of these videos were removed before any video views.

    - **84.5%** were removed within 24 hours of being posted.

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 does not allow users to post content containing misinformation (or disinformation) 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.

- We rely on independent fact-checking partners and our database of previously fact-checked claims to help assess the accuracy of content. Content is also ineligible for the FYF if it contains general conspiracy theories or unverified information related to emergencies. To be cautious, content that warrants fact-checking is also temporarily ineligible for the FYF while it is undergoing review.

- Violating our Community Guidelines can lead to enforcement actions on accounts (including removal). We do not allow the trade of services that attempt to artificially increase engagement or deceive TikTok's recommendation system. If we become aware of accounts or content with inauthentically inflated metrics, we will remove the associated fake followers or likes. Content is ineligible for the FYF if it seeks to trick or manipulate others to increase followers, likes, or views.

- As part of updates to our Community Guidelines that came into effect April 2023, TikTok advanced 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.

- We proactively identify and remove covert influence operations and publicly disclose information about specific attacks in our Community Guidelines Enforcement reports.

- We also empower users to make informed decisions about the content and information they consume on TikTok, providing publicly available account information (no. 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.

- In January 2023, we expanded our state-controlled media label program to over 40 markets across multiple regions - ensuring our users have accurate, transparent, and actionable context when they engage with content from media accounts that may present the viewpoint of a government.

- 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 our website by choosing a reason why they think the content might violate our Community Guidelines.

- 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 promotion, sale, or solicitation of or facilitation of access to products or services that might be or are considered deceptive, misleading, or unlawful such as: unwarranted claims, misinformation, including pricing/discount or promotion information inconsistency, missing T&C or privacy policy pages, or any of such. We also require 18+ targeting for certain categories of products and service and proper disclaimers must be included when applicable.

- TikTok has a range of online resources, in-app PSAs and hubs developed in partnership with reputable third-parties focussing on a range of important topics including COVID-19, elections, and mental health. In preparation for the General Election on 14 October 2023, TikTok has partnered with Elections NZ to launch in-app PSAs (search guide interventions) that will prompt users to refer to official resources when engaging with election related information.

- In October 2022, TikTok engaged the Australian Associated Press to create and deliver targeted digital media literacy education to a small group of content creators, including from New Zealand. The workshops empowered 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 had large followings, with the project focussed on having a wide impact in educating audiences on how to identify mis- and disinformation.

- In Q3 2023, TikTok is also convening an internal academic/expert roundtable discussing potential online safety challenges, including online disinformation, and recommendations.

Relevant *Global* Community Guidelines Enforcement Statistics:

- In Q1 2023, less than **1%** of all videos uploaded globally were removed for violating our Community Guidelines.

- In Q1 2023, **1.0%** of total videos removed violated our integrity and authenticity policies.
    - **94.8%** were proactively removed before any video reports.
    - **72.8%** of these videos were removed before any video views.
    - **76.6%** were removed within 24 hours of being posted.

- In Q1 2023, TikTok removed **more than** 51 million fake accounts, 531 million fake likes, 1.2 trillion fake followers globally.

Relevant *New Zealand* Community Guidelines Enforcement Statistics:

- In Q4 2022, nearly 160,000 videos were removed in New Zealand for violating our Community Guidelines.
    - **95.8%** were proactively removed before any video reports.
    - **78.0%** of these videos were removed before any video views.
    - **88.3%** were removed within 24 hours of being posted.

- In Q1 2023, nearly 120,000 videos were removed in New Zealand for violating our Community Guidelines.

    o **95.1%** were proactively removed before any video reports.

    o **66.1%** of these videos were removed before any video views.

    o **84.5%** were removed within 24 hours of being posted.

Regarding **Measure 31:**

- Paid political advertising, including sponsored political content, is not allowed on TikTok. For more information regarding TikTok's advertising policies, please see [here](#).

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]

- As explained in our baseline report, TikTok presents 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. It 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. Content categories that are ineligible for the For You feed (FYF) are detailed in our Community Guidelines.

- In December 2022, we launched the 'Why this video?' tool to help people understand why a particular video has been recommended to them.

- The For You feed can 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 – e.g. a vegetarian who wants to see fewer meat recipes.

- We further empower users to make informed decisions about the content and information they consume on TikTok, providing publicly available account information (no. 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.

- In January 2023, we expanded our state-controlled media label program to over 40 markets across multiple regions - ensuring our users have accurate, transparent, and actionable context when they engage with content from media accounts that may present the viewpoint of a government.

- TikTok also partners with third party fact checking organisations to detect and mitigate the spread of misinformation and disinformation in Oceania markets; leveraging the insights of IFCN-certified fact-checking experts.

- In October 2022, TikTok engaged the Australian Associated Press to create and deliver targeted digital media literacy education to a small group of content creators, including from New Zealand. The workshops empowered 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 had large followings, with the project focussed on having a wide impact in educating audiences on how to identify mis- and disinformation.

- TikTok also has a range of online resources, in-app PSAs and hubs developed in partnership with reputable third-parties focussing on a range of important topics including COVID-19, elections, and mental health that provide users access to reliable information. In preparation for the General Election on 14 October 2023, TikTok has partnered with Elections NZ to launch in-app PSAs (search guide

interventions) that will prompt users to refer to official resources when engaging with election related information.

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.

Regarding Measure 37:

- TikTok presents 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. It 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. Content categories that are in ineligible for the For You feed (FYF) are detailed in our Community Guidelines.

- In December 2022, we launched the 'Why this video?' tool to help people understand why a particular video has been recommended to them.

- The feed can 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 – e.g. a vegetarian who wants to see fewer meat recipes.

- 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 violate our Community Guidelines. All these actions contribute to future recommendations in the For You feed.

- In March 2023, TikTok launched a "Refresh" feature allowing users 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 or not
    - Choosing who can see their following list i.e. people they follow

- Our Family Pairing feature allows give parents and guardians more control over their teen's experience on TikTok. Since 1 March 2023, as part of updates to Family Pairing, every account belonging to a user below age 18 has automatically set on their account a 60-minute daily screen time limit.
- In addition to these new features to Family Pairing, everyone will soon be able to set their own customised screen time limits for each day of the week and set a schedule to mute notifications. In addition, we're rolling out a sleep reminder to help people more easily plan when they want to be offline at night. People can set a time, and when it's reached, a pop-up will remind them it's time to log off.

Regarding **Measure 38**:

- TikTok enforces strict advertising policies that go beyond industry standards to protect all users on our platform.
- 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: gambling, tobacco, alcohol, 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 minors 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 company'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.

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'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 Wellbeing, Mental and Behavioural Health and Integrity and Authenticity.

- 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. This occurred as recently as April 2023.

- TikTok also has a range of online resources, in-app PSAs and hubs developed in partnership with reputable third-parties focussing on a range of important topics including COVID-19, elections, and mental health.

- 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.

- We also publish quarterly Community Guideline Enforcement Reports, with additional Transparency Reports every six months on Government and Law Enforcement requests.

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 uses a combination of AI and human moderation to identify, review, and action content that violates our Community Guidelines and Terms of Service. 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 and has provided our first annual compliance report for 2023 as required.

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]

- As explained in our baseline report, TIkTok is 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 recent research, partnerships and multi-stakeholder programs we have supported include:

- In July 2022, TikTok made several commitments around transparency with the research community, including to build a Research API. This new tool has now launched to academic researchers in the US and Europe, and further expansions are planned, including for the Asia Pacific. The Research API supports research on a range of topics, including online safety, by providing access to public data on TikTok content and accounts. To date, the most common research topics in applications for API access include consumer trends, mental health, and misinformation. TikTok will also be presenting the Research API at the upcoming Trust & Safety Research Conference hosted by Stanford University.

- For Netsafety Week 2022, TikTok partnered with Netsafe and Professor Amanda Third to host a webinar to provide an educational session for parents in NZ - this included a strong focus on protecting children online, including education about our family pairing feature, age restriction information and community controls. Amanda is an expert in user-centred, participatory research, and her work investigates children's and young people's technology practices.

- In October 2022, TikTok engaged the Australian Associated Press to create and deliver targeted digital media literacy education to a small group of content creators, including from New Zealand. The workshops empowered 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 had large followings, with the project focussed on having a wide impact in educating audiences on how to identify mis- and disinformation.

- In November 2022, TikTok partnered with leading academic experts, NGOs and regulators on the TikTok Creator Safety Summit, which trained top creators, including from New Zealand, in the online prevention of Bullying, Harassment and Hateful Behaviour.

- In October 2023, TikTok is co-sponsoring the 'Breaking Glass Ceilings and Building Bridges: a trans-Tasman conference exploring the human experience in online spaces and places' conference organised by Netsafe New Zealand in association with the Australian eSafety Commissioner. The conference brings 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 focussing on online harm and identity, including:

  o *Legislative and policy responses*

  o *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]

- As founding signatories 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) 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 and has provided our first annual compliance report for 2023 as required.