

Aotearoa New Zealand Code of Practice for Online Safety and Harms

Google Annual Report

Signatory: Google

Relevant Products/Service: YouTube

Submitted: October 2024



We are pleased to provide our Annual Report under the Aotearoa New Zealand Code of Practice for Online Safety and Harms for the reporting period of 1 July 2023 - 30 June 2024. This Report builds on the information provided under our [Baseline Report of 2022](#) and subsequent report(s) [here](#). It is arranged following the Code's key sections 4.1 to 4.4.

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 1.3.

YouTube is an open video platform where anyone can upload a video and share it with the world. With this openness comes incredible opportunities, as well as challenges – which is why we're always working to balance creative expression with our responsibility to protect the community from harmful content.

Our previous reports provided a detailed outline of YouTube's approach to content moderation. The following is a summary of that information with updated metrics from our most recent [YouTube Community Guidelines Enforcement report](#):

- YouTube's approach to combating harmful content involves removing content that violates YouTube's policies as quickly as possible, surfacing high quality information in ranking and recommendations, and rewarding trusted, eligible creators and artists:
 - **A focus on responsible recommendations:** Recommendations play an important role in how YouTube maintains a responsible platform. Recommendations connect viewers to high-quality information and minimise the chances to see problematic content. Not all content is recommended. Recommendations also complement the work done by our robust [Community Guidelines](#) that define what is and isn't allowed on YouTube.
 - **Raising quality content:** We take the additional step of recommending authoritative videos to viewers on topics such as news, medical, and scientific information. We rely on human evaluators, trained using publicly available guidelines, who assess the quality of information in each channel and video. To decide if a video is authoritative, evaluators look at factors like the expertise and reputation of the speaker or channel, the main topic of the video, and whether the content delivers on its promise or achieves its goal. The more authoritative a video, the more it's promoted in recommendations.
- We have robust mechanisms to monitor compliance with, and to enforce, our policies. We rely on a mix of human and technological intervention:
 - we encourage all users to [report](#) content that violates our [Community Guidelines](#);



- we have established the [YouTube Priority Flagger program](#), by which government agencies and NGOs, such as Netsafe, can notify us of content that violates our [Community Guidelines](#); and
- we have developed machine learning classifiers to automatically and quickly identify and remove potentially violative content.

Content that violates our policies will be removed, and repeated violations or a single case of severe abuse may result in channel terminations.

YouTube Community Guidelines enforcement globally, July 2023 - June 2024

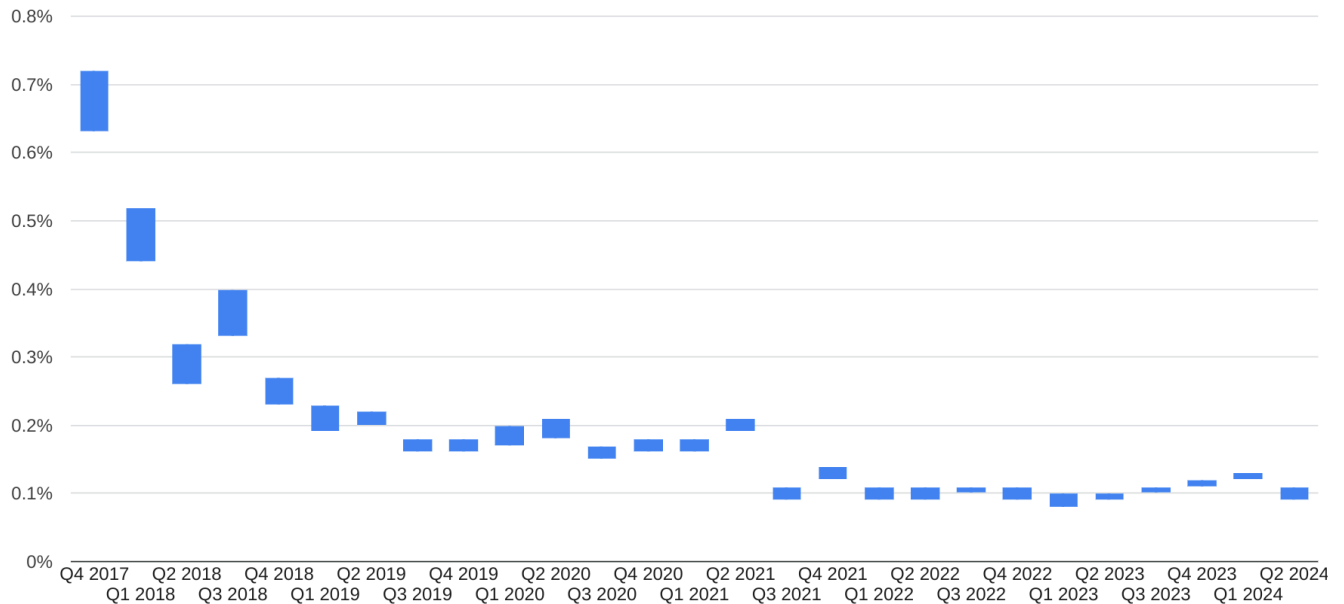
	Q3 2023	Q4 2023	Q1 2024	Q2 2024
YouTube channels were removed globally for violating our Community Guidelines	10,501,720	20,592,341	15,799,880	3,260,974
YouTube videos were removed globally for violating our Community Guidelines	8,115,659	9,012,232	8,295,304	8,497,876
YouTube videos removed globally were first detected by automated flagging	7,752,654	8,656,644	7,996,564	8,198,119
Percentage of YouTube videos removed globally before they received more than 10 views	76.4%	77.9%	82.7%	84.3%
Comments were removed globally for violating our Community Guidelines	842,831,976	1,197,687,868	1,443,821,162	1,372,493,981
Percentage of removed comments globally first detected by automated flagging	99.5%	99.6%	99.7%	99.6%

- YouTube strives to remove content that violates our [Community Guidelines](#) before users are exposed to this content. We use a metric called the [Violative View Rate \(VVR\)](#) - an estimate of the proportion of video views that violate our Community Guidelines in a given quarter (excluding spam) - to measure our progress on removing violative videos.
 - In **Q3 2023** (July - Sept 2023), VVR was 0.10 - 0.11% (i.e., out of every 10,000 views on YouTube, 10-11 were of violative content).
 - In **Q4 2023** (Oct - Dec 2023) VVR was 0.11 - 0.12% (i.e., out of every 10,000 views on YouTube, 11-12 were of violative content).
 - In **Q1 2024** (Jan - Mar 2024), VVR was 0.12 - 0.13% (i.e., out of every 10,000 views on YouTube, 12-13 were of violative content).



- In **Q2 2024** (Apr - Jun 2024), VVR was 0.09 - 0.11% (i.e., out of every 10,000 views on YouTube, 9-11 were of violative content).

The VVR is [published quarterly](#) as part of our [YouTube Community Guidelines Enforcement report](#). Further information on the VVR methodology can be found [here](#).



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

Google is committed to fighting child sexual exploitation and abuse online and preventing our platforms from being used to spread child sexual abuse material (CSAM).

Policies

YouTube's [Community Guidelines](#) prohibit sexually explicit content featuring minors and content that sexually exploits minors.

- Our [child safety policy](#) prohibits content that endangers the emotional and physical well-being of minors. This includes content that sexually exploits minors, that highlights harmful or dangerous acts involving minors, and that causes emotional distress to minors.
- Our [nudity and sexual content policy](#) prohibits content that is meant to be sexually gratifying.

Enforcement

Global Removals

From July 2023 to June 2024, the following actions were taken globally:



	Q3 2023	Q4 2023	Q1 2024	Q2 2024
YouTube channels were removed for violating our nudity and sexual content policy	420,428	917,941	159,119	156,802
YouTube channels were removed for violating our child safety policy	93,443	111,291	138,766	164,435
YouTube videos were removed for violating our nudity and sexual content policy	609,517	493,168	470,262	419,560
YouTube videos were removed for violating our child safety policy	2,991,727	2,921,829	3,598,821	5,046,178

New Zealand Removals

From July 2023 to June 2024, the following actions were taken in respect of YouTube videos uploaded from IP addresses in New Zealand:

	Q3 2023	Q4 2023	Q1 2024	Q2 2024
YouTube videos uploaded from IP addresses in New Zealand were removed for violating our child safety policy	2,305	2,500	3,428	3,340
YouTube videos uploaded from IP addresses in New Zealand were removed for violating our nudity and sexual content policy	346	293	273	333

Product, Policy and Partnership Updates

- YouTube has continued to report videos containing CSAM or where a user solicits CSAM through comments or other communications to the National Centre for Missing and Exploited Children (NCMEC), which liaises with law enforcement agencies around the world, including in New Zealand:

From July 2023 to June 2024, the following actions were taken globally:

	H2 2023	H1 2024
Total pieces of content reported to NCMEC by YouTube	265,371	320,498
Cyber Tipline reports to NCMEC by YouTube	225,440	280,478



Further child safety information and metrics can be found in both [YouTube's Community Guidelines Enforcement report](#) and Google's [Combating Child Sexual Abuse Material Transparency report](#).

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

YouTube's mission is to give everyone a voice and show them the world. We believe that everyone deserves to have a voice, and that the world is a better place when we listen, share and build community through our stories. Harassment hurts our community by making people less inclined to share their opinions and engage with each other.

Policies

Our [harassment and cyberbullying policy](#) prohibits content that targets someone with prolonged insults or slurs based on their physical traits or [protected group status](#). This also includes harmful behaviour such as threats or doxxing (i.e., publishing private or identifying information about an individual online, typically with malicious intent).

Enforcement

Global Removals

From July 2023 to June 2024, the following actions were taken globally:

	Q3 2023	Q4 2023	Q1 2024	Q2 2024
YouTube channels were removed for harassment and cyberbullying	39,419	30,345	26,908	30,509
YouTube videos were removed for harassment and cyberbullying	513,854	444,342	448,313	504,265

New Zealand Removals

In July 2023 to June 2024, the following actions were taken in respect of YouTube videos uploaded from IP addresses in New Zealand:



	Q3 2023	Q4 2023	Q1 2024	Q2 2024
YouTube videos uploaded from IP addresses in New Zealand were removed for harassment and cyberbullying	552	407	316	385

Product, Policy and Partnership Updates

- **Update to Harassment & cyberbullying enforcement:** In January 2024, YouTube began striking content that realistically simulates deceased minors or victims of deadly or well-documented major violent events describing their death or violence experienced based on updates to the [Harassment & cyberbullying policy](#).
- **Continued investment in creator safety:** To address creator concerns about hurtful and hateful comments, our teams continue to invest in systems improvements while also launching new features designed to help creators more easily moderate their comments:
 - Creators have an [optional setting](#) that will catch even more potentially inappropriate or spam comments by selecting “Increase Strictness” in their comment settings. These comments are [held for review](#) in YouTube Studio, where creators can choose to approve, remove or report them.
 - We also made an update to YouTube Studio so that potentially offensive comments are now placed in a separate, hidden section at the bottom of the held for review tab. This way, creators can choose to ignore them completely and leave them unreviewed if they prefer.
- In collaboration with creators and third-party experts like [ConnectSafely](#), [The Family Online Safety Institute](#) and the [National Cybersecurity Alliance](#), we collected in depth information and tips on topics like how to stay safe when starting out as a new creator, what to do as your channels are growing and how to navigate experiencing things like bullying, trolling, account hijacking and more. Creators continue to have access to all this information in our [Creator Safety Center](#).
- In November 2023, YouTube updated its [Community Posts Policy enforcement](#). While YouTube has always removed Community posts that violate its policies, starting in November 2023, these posts may also result in a Community Guidelines strike applied to the channel.
- Beginning in March 2024, YouTube, in some cases, may automatically set some videos to private when it detects that videos may have been uploaded and published without the channel owner’s knowledge or permission. This update is part of our efforts to keep creators and their channels safe against hijacking. As an added precaution, the channel owner will be signed out of their account and notified by email.



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

Policies

Hate speech is not allowed on YouTube.

- Our [hate speech policy](#) prohibits content that promotes violence or hatred against individuals or groups based on any of the following attributes:
 - age
 - caste
 - disability
 - ethnicity
 - gender identity and expression
 - nationality
 - race
 - immigration status
 - religion
 - sex/gender
 - sexual orientation
 - victims of a major violent event and their kin, or
 - veteran status.

This means we don't allow content that dehumanises individuals or groups with these attributes, claims they are physically or mentally inferior, or praises or glorifies violence against them. We also don't allow use of stereotypes that incite or promote hatred based on these attributes, or racial, ethnic, religious, or other slurs where the primary purpose is to promote hatred.

Sometimes content directed against an individual may instead be covered by our policies against [harassment](#) and [violence](#), while content that praises or glorifies terrorist or criminal figures or organisations is covered by our policies against [violent criminal organisations](#). Reviewers evaluate flagged content against all of our [Community Guidelines](#) and policies.

Enforcement

Global Removals

From July 2023 to June 2024, the following actions were taken globally:

	Q3 2023	Q4 2023	Q1 2024	Q2 2024
YouTube channels were removed for being hateful or abusive	26,130	96,008	37,473	34,020
YouTube videos were removed for being hateful or abusive	186,999	181,409	157,831	163,000



New Zealand Removals

From July 2023 to June 2024, the following actions were taken in respect of YouTube videos uploaded from IP addresses in New Zealand:

	Q3 2023	Q4 2023	Q1 2024	Q2 2024
YouTube videos uploaded from IP addresses in New Zealand were removed for being hateful or abusive	530	490	363	328

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

Policies

Content encouraging others to commit violent acts is not permitted on YouTube:

- Our [violent and graphic content policy](#) prohibits content that incites others to commit violent acts against individuals or a defined group of people.
- Content intended to praise, promote or aid violent criminal organisations is prohibited under our [violent criminal organisations policy](#). This includes content produced by violent criminal or terrorist organisations, content praising or memorialising prominent terrorist or criminal figures in order to encourage others to carry out acts of violence and content aimed at recruiting new members to violent criminal or terrorist organisations, or content that glorifies or promotes violent tragedies.
- Our [harmful or dangerous content policy](#) does not allow weapons content that includes instructions that show or tell viewers how to perform activities that are meant to kill or severely harm others, or giving instructions to make explosive devices or compounds meant to injure or kill others.
- Our hate speech policy prohibits the incitement of violence against individuals or groups based on protected attributes (see Outcome 3 above).

Reviewers evaluate flagged content against all of our [Community Guidelines](#) and policies.

Enforcement

Global Removals

From July 2023 to June 2024, the following actions were taken globally:



	Q3 2023	Q4 2023	Q1 2024	Q2 2024
YouTube channels were removed for the promotion of violence and violent extremism	32,339	42,885	36,806	29,153
YouTube videos were removed for the promotion of violence and violent extremism	163,417	365,112	311,036	252,470

New Zealand Removals

From July 2023 to June 2024, the following actions were taken in respect of YouTube videos uploaded from IP addresses in New Zealand:

	Q3 2023	Q4 2023	Q1 2024	Q2 2024
YouTube videos uploaded from IP addresses in New Zealand were removed for the promotion of violence and violent extremism	70	105	100	82

Product, Policy and Partnership Updates

- YouTube has continued to work across industry and multi stakeholder groups like the Christchurch Call and Global Internet Forum to Counter Terrorism (GIFCT) to combat terrorist and violent extremist content on the Internet. Since our last report, GIFCT's [Content Incident Protocol](#) has been activated in response to the January 2024 event in Perry, Iowa, and the January 2024 event in Levittown, Pennsylvania.
- In March 2024, YouTube's [policy on Harmful & Dangerous content](#) was updated to include a stricter stance on audience disclaimers and updated guidelines to better assess the risk of potential harm of the act portrayed.
- Starting in June 2024, content showing how to remove certain safety devices will be prohibited based on YouTube's update to the [Firearms policy](#). Content showing the use of homemade firearms, automatic firearms, and certain firearm accessories will be age restricted. We will also expand enforcement on content sharing links where you can purchase firearms, ammo, or certain accessories.

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



Policies

Gory and violent content intended to shock or disgust viewers is not allowed on our platform.

- Our [violent or graphic content policy](#) prohibits violent and graphic content as well as animal abuse content, and extends to dramatised or fictional content where the viewer is not given enough content to understand that the footage is dramatised or fictional.

Violent and graphic content may also be captured under our [violent criminal organisations policy](#) and [harmful or dangerous content policy](#).

Reviewers evaluate flagged content against all of our [Community Guidelines](#) and policies. We may make exceptions for content that has a clear educational, documentary, scientific, or artistic purpose. For example, we may allow news outlets to publish footage from a war zone if educational or documentary intent is clear, but we may still place it behind an interstitial warning to users that the content is graphic so users know that the content may not be suitable for all audiences. We think this policy is critical to striking the right balance with free expression, and serves important societal purposes. Note that [content that does not get exceptions](#), even if there's educational, documentary, scientific, or artistic context provided, includes, but is not limited to, the following:

- Violent physical sexual assaults (video, still imagery, or audio).
- Footage filmed by the perpetrator during a deadly or major violent event, in which weapons, violence, or injured victims are visible or audible.

Enforcement

Global Removals

From July 2023 to June 2024, the following actions were taken globally:

	Q3 2023	Q4 2023	Q1 2024	Q2 2024
YouTube channels were removed for violent or graphic content	14,333	16,846	22,516	15,686
YouTube videos were removed for violent or graphic content	899,221	672,029	785,851	646,591

New Zealand Removals

From July 2023 to June 2024, the following actions were taken in respect of YouTube videos uploaded from IP addresses in New Zealand:

	Q3 2023	Q4 2023	Q1 2024	Q2 2024
--	---------	---------	---------	---------



YouTube videos uploaded from IP addresses in New Zealand were removed for violent or graphic content	257	277	326	263
--	-----	-----	-----	-----

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

Policies

Certain types of misleading or deceptive content with serious risk of egregious harm is not allowed on our platform. YouTube’s misinformation policies, detailed below, can be accessed via this [landing page](#):

- **Misinformation policy:** Certain types of misleading or deceptive content with serious risk of egregious harm are not allowed on YouTube. This includes certain types of misinformation that can cause real-world harm, certain types of technically manipulated content, or content interfering with democratic processes. This type of prohibited content includes suppressing census participation, manipulated content, or misattributed content.
- **Election misinformation policy:** Aligned with the misinformation policy, certain types of misleading or deceptive content with serious risk of egregious harm are not allowed on YouTube. This includes misinformation that can cause real-world harm, like certain types of technically manipulated content, and content interfering with free and fair democratic election processes. This prohibited content includes voter suppression, candidate eligibility, elections integrity, and incitement to interfere with democratic processes.
- **Medical misinformation policy:** YouTube does not allow content that poses a serious risk of egregious harm by spreading medical misinformation that contradicts local health authorities’ or World Health Organization (WHO)’s guidance about specific health conditions and substances. This prohibited content includes prevention misinformation, treatment misinformation, and denial misinformation.

Enforcement

Global Removals

From July 2023 to June 2024, the following actions were taken globally:

	Q3 2023	Q4 2023	Q1 2024	Q2 2024
YouTube channels were removed for misinformation	181,845	183,257	139,693	148,408



YouTube videos were removed for misinformation	51,920	171,828	117,065	104,200
--	--------	---------	---------	---------

New Zealand Removals

From July 2023 to June 2024, the following actions were taken in respect of YouTube videos uploaded from IP addresses in New Zealand:

	Q3 2023	Q4 2023	Q1 2024	Q2 2024
YouTube videos uploaded from IP addresses in New Zealand were removed for misinformation	105	320	215	171

Product, Policy and Partnership Updates

- **Responsible AI innovation:** In November 2023, YouTube published its [approach to responsible AI innovation](#), which outlined its plan to require that creators disclose when the content they are uploading is made with altered or synthetic media, including generative AI, and is realistic - that is, a viewer could easily mistake what is being shown with a real person, place or event. YouTube announced that, tied to these disclosures, it will begin applying labels to content indicating that some of the content was altered or synthetic, as well as a more prominent label for certain types of content about sensitive topics. YouTube officially rolled out the disclosure requirement and tools for creators in Q1 2024.
- **Updated approach to medical misinformation:** Content that promotes harmful or ineffective cancer treatments will now be removed. Starting September 2023, violations to this policy can result in a channel strike. Cancer treatment misinformation includes content that promotes unproven treatments (in place of approved care) or as a guaranteed cure. It also includes promotion of any treatments that have been specifically deemed harmful by health authorities. [Learn more.](#)
- **Hit Pause, YouTube’s global media literacy campaign:** Throughout the reporting period, YouTube has continued to launch additional [Hit Pause](#) content, an educational program seeking to teach viewers critical media literacy skills. In October 2023, YouTube released a new Hit Pause campaign, titled “Evaluate & Share” consisting of 3 videos that aim to educate users on how to evaluate the credibility of sources and share information responsibly.
- **Grant to the International Fact Checking Network (IFCN):** Google and YouTube take our role in helping to fight misinformation seriously and we continually invest in products, programs and partnerships to help people access high-quality information. In November 2022, Google and YouTube announced a USD 13.2 million grant to the [International Fact-Checking Network](#) (IFCN) to launch a new Global Fact Check Fund to support their network of 135 fact-checking organisations from 65 countries covering over 80 languages. Building on our



previous work to address misinformation, this is Google and YouTube's single largest grant in fact-checking. For regular updates on application windows for each grant round, announcements about which organisations were independently selected, and information about grant distribution, visit the IFCN website [here](#).

- **Access to information during an Election Period:** In the build-up to elections, people needed helpful and authoritative information to help them navigate the electoral process. For the New Zealand election in October 2023, we made electoral information from the [Election Commission](#) and other authoritative sources - such as how to vote and where to vote - easily discoverable on YouTube through public service announcements (PSAs) on our homepage. See [here](#) for more information.
- At a local level, we are also continuing to work with industry and other stakeholders to respond to evolving harms arising from misinformation:
 - Partnered with AAP FactCheck to increase their capacity to fact check, audience reach and analyse misinformation trends. An example is highlighted in this [article](#) published on 17 June, 2024.
 - [Partnered with Squiz Kids'](#) media literacy module, "Newshounds" to launch its plug and play resources for teachers, children and their parents in New Zealand and Australia.
 - Continue to sponsor an [online tracking tool](#) for journalists to understand, monitor and report on local candidates ahead of New Zealand's local elections.

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

Policies

Our Community Guidelines include tough policies against users that deliberately try to deceive or mislead people.

- **[Impersonation policy:](#)** This policy states that content intended to impersonate a person or channel is not allowed on YouTube. YouTube also enforces trademark holder rights. When a channel, or content in the channel, causes confusion about the source of goods and services advertised, it may not be allowed.
- **[Fake engagement policy:](#)** YouTube does not allow anything that artificially increases the number of views, likes, comments, or other metrics either by using automatic systems or serving up videos to unsuspecting viewers. Content and channels that do not follow this policy may be terminated and removed from YouTube.
- **[Spam, deceptive practices, and scam policies:](#)** YouTube does not allow spam, scams, or other deceptive practices that take advantage of the YouTube community. We also do not allow content where the main purpose is to trick users into leaving YouTube for another site.



Enforcement

Global Removals

From July 2023 to June 2024, the following actions were taken globally:

	Q3 2023	Q4 2023	Q1 2024	Q2 2024
YouTube channels were removed for spam, misleading and scams	9,625,070	19,101,658	15,163,210	2,623,270
YouTube videos were removed for spam, misleading and scams	362,453	146,932	92,167	108,267

New Zealand Removals

From July 2023 to June 2024, the following actions were taken in respect of YouTube videos uploaded from IP addresses in New Zealand:

	Q3 2023	Q4 2023	Q1 2024	Q2 2024
YouTube videos uploaded from IP addresses in New Zealand were removed for spam, misleading and scams	583	114	44	207

Product, Policy and Partnership Updates

- Google Advertising launched an API available globally in the [Ads Transparency Center](#). The API became available starting in Q3 2023.



Ads transparency for a safe and open internet

Find active ads published through Google

Any time ▾ Shown in New Zealand ▾ All formats ▾

🔍 All topics ▾ Search by advertiser or website name

More information can be found [here](#).

- **The latest on our efforts to counter government-backed attacks from the Threat Analysis Group:** The Threat Analysis Group (TAG) tracks actors involved in disinformation campaigns. The actions taken against coordinated influence operation campaigns on our platforms, during the reporting period, are disclosed in the Quarterly Bulletin (published on our [Threat Analysis Group blog](#)).

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 behaviour 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



Given the scale of content uploaded to YouTube, finding what you need would be nearly impossible without some help sorting through all the videos. YouTube's search ranking system does just that by sorting through a vast number of videos to find the most relevant and useful results to your search query, and presenting them in a way that helps you find what you're looking for.

Recommendations play an important role in how YouTube maintains a responsible platform. Recommendations connect viewers to high-quality information and minimise the risk of seeing problematic content. Not all content is recommended. Recommendations also complement the work done by our robust [Community Guidelines](#) that define what is and isn't allowed on YouTube.

Context is critical when evaluating information, so we give you context alongside certain search results and videos to help you make your own decisions about the content that you find on YouTube. To do this, we highlight text-based information about certain events, topics, and publishers from third-party sources using information panels across YouTube.

At the same time, YouTube has and will continue to offer users a range of tools to optimise their experience on the platform and decide what personal information is used to influence recommendations. For example, signed-in users can clear their search history, remove individual search entries from search suggestions, or pause search history using the YouTube history settings. Finally, in-product controls enable users to remove recommended content from their home pages and watch next.

Further examples of some tools and features we have created to provide users with the opportunity to make informed choices about the content they encounter are detailed in our previous reports, including:

- [YouTube Kids](#) and [YouTube supervised experience](#) for families and younger users
 - YouTube regularly updates its [YouTube for Families Help Centre](#) to share safety resources for families with children or teens who use YouTube Kids or Supervised Experiences. This Help Centre page is accessible to everyone and provides information about YouTube's products and features that parents and caregivers can use to guide their kids' YouTube journey and choose the best YouTube experience for their family.
 - YouTube has a [family guide](#) for supervised experiences that outlines tips and tricks that parents and caregivers can leverage to help build healthy digital habits.
 - [Verification badges](#) to inform users where we have verified a channel as the official channel of a creator, artist, company or public figure and help users distinguish official channels from other channels with a similar name on YouTube.
- Google uses a combination of people and machine learning tools to detect and flag potentially age inappropriate content for minors, which is then [age-restricted](#). YouTube also provides an option for creators to age-restrict their own videos. This type of age restriction is self-imposed and is not a result of a review by YouTube. When a video is age-restricted, viewers must be signed in and 18 years of age or older to view it. Age-restricted videos are not



shown in certain sections of YouTube and may also have limited or no ads. If YouTube decides that a video should be age restricted, a permanent age restriction will be applied. This restriction happens even if the creator also age restricts the video.

Product, Policy and Partnership Updates

- In November 2023, [YouTube announced](#) our [safeguards reducing repeated recommendations](#) of certain content to specifically support the well-being and mental health of young people. For example, while some types of content may seem innocuous in a single view, experts have found that repeatedly viewing some types of content can be problematic for young people, such as content that displays social aggression, compares physical features or idealises some body types, fitness levels or weight over others.
- **Google Ads Update to Political Content Policy:** In mid-November 2023, Google Ads [updated its Political content policy](#) to include disclosure requirements for synthetic content. This policy required that all verified election advertisers in regions where verification is required must prominently disclose when their ads contain synthetic content that authentically depicts real or realistic-looking people or events. This disclosure must be clear and conspicuous, and must be placed in a location where it is likely to be noticed by users. This policy will apply to image, video and audio content.

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

As detailed in our response to Outcome 8, YouTube offers tools to enable users to control the content they see on our platform through supervised accounts for children, and by managing their watch history. In addition to these tools, a user can also subscribe to channels they like to see more content from those channels. Once a user subscribes to a channel, any new videos it publishes will show up in the Subscriptions feed.

Additionally, YouTube enables users to turn personalised ads on or off through [My Ad Center](#). If they're turned on, we also allow you to turn on or off particular categories of ads (like apparel, banking, etc). To better protect children, personalised ads, remarketing, and other personalised targeting features are [prohibited](#) on YouTube for:

- Google Accounts managed by Family Link for children under the age of 13 (supervised accounts); and
- Content set as [made for kids](#).

Contextual ads can be served on YouTube for supervised accounts and on content set as made for kids. These ads are based on factors such as:

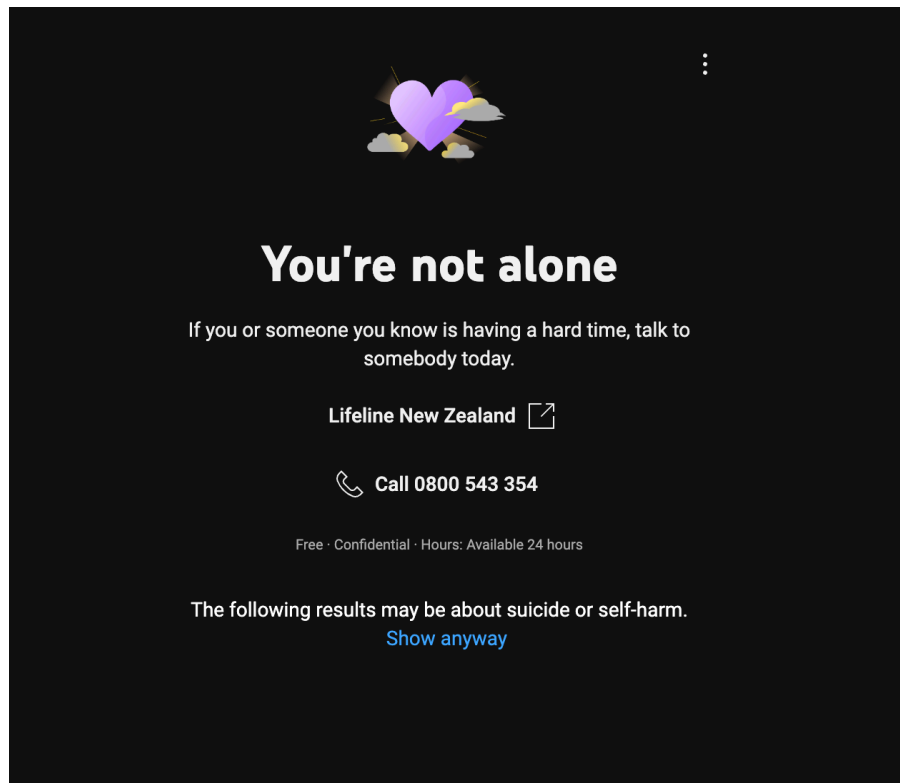
- The content being viewed.

- The viewer's current search.
- The viewer's general location (such as city or state).

Ads must follow the [made for kids ad policy](#) to be eligible to appear on YouTube for supervised accounts and on content set as made for kids.

Viewers of "made for kids" content may see an ad bumper before and after a video ad is shown. This bumper helps alert them when an advertisement is starting and ending. If viewers have a [YouTube Premium family plan](#), their children are eligible for ad-free content and other shared benefits of membership.

YouTube also recently [expanded crisis](#) resource panels into a new full-page experience that better helps viewers pause for a moment. This full-page experience allows viewers to more prominently see resources for the third-party crisis hotlines run by locally based organisations. An example of the full-page experience of the crisis resource panel in New Zealand is shown in the screenshot below:



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

The success of our business is built on providing trusted products and services, and transparency about how we organise content is essential to that trust.

Our policies work best when users are aware of the rules and understand how we enforce them. That is why we work to make this information clear and easily available to all. Our previous reports provided details on some of our efforts in this regard, including:

- Our comprehensive [help centre](#) with detailed information about policies—including our [Community Guidelines](#) and Advertiser-Friendly Guidelines.
- [How YouTube Works](#) — a website designed to answer the questions we most often receive about what we’re doing to foster a responsible platform for our community, and explain our products and policies in detail. How YouTube Works provides an in-depth look at our products and settings, such as YouTube Search, Recommendations, privacy controls, and Ad Settings, showing how they help people have the best possible experience while they’re using YouTube.
- The [YouTube Official Blog](#) also provides further detail on how YouTube works, including the development and evolution of our policies and products.
- Our [Transparency Center](#), a central hub for quickly and easily learning more about Google’s product policies, including YouTube’s. The Transparency Center collects existing resources and policies, and was designed with you in mind, providing easy access to information on our policies, how we create and enforce them, and much more, including:
 - Our policy development process
 - Policies by product or service
 - Reporting & appeal tools
 - Transparency reports
 - Google’s principles for privacy and AI

Further information can be found [here](#).

Updates



- YouTube continues to assess, evaluate, and update its policies on a regular basis, the latest updated policies, including [Community Guidelines](#), can be found [here](#).
- In August 2023, YouTube announced changes to its [Community Guidelines warnings](#). After the creator's first Community Guidelines violation, they will typically get a warning with no penalty to their channel. They will now have the chance to take policy training to allow the warning to expire after 90 days.

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

Since Google launched its first Transparency Report in 2010, we've been sharing data that sheds light on how the policies and actions of governments and corporations affect privacy, security, and access to information online. Our previous reports provide detail on Google and YouTube's transparency reporting efforts:

- Our [YouTube Community Guidelines Enforcement report](#) provides a quarterly update on the work we do to enforce our policies on YouTube. The report offers data on global video, channel, and comment removals for violating our policies; appeals and reinstatements; and human and machine flagging.
- Our [Google Transparency Report website](#) is a centralised hub for transparency reporting on key topics including child safety, copyright, political advertising, government requests to remove content, and government requests for user information.
- Our [Transparency Centre](#) is a central hub for quickly and easily learning more about Google's product policies, including YouTube's. The Transparency Center collects existing resources and policies, and was designed with you in mind, providing easy access to information on our policies, how we create and enforce them, and much more. See outcome 10 for more detail.
- The annual [Ads Safety Report](#) provides updates on policy enforcement in Google Ads, including ads shown on YouTube.
- Our Threat Analysis Group Quarterly Bulletin (published on our [Threat Analysis Group blog](#)) discloses actions we have taken against coordinated influence operation campaigns on our platforms.
- Our website [How YouTube Works](#) provides detailed information about how we use algorithms to rank and recommend content.

As mentioned, Google makes available data on Government requests to remove content, for instance where a Government body claims that content violates local law or a government requests that we review content to determine if it violates our own product [Community Guidelines](#) and



content policies. Data on removal requests for New Zealand can be found [here](#).

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.

YouTube continues to support global and local research efforts, such as our partnership with Squiz Kids' media literacy module; "Newshounds", which is a plug and play resource for teachers, children and their parents in NZ and AU; and with Te Rito to fund and provide a training camp for Māori and Pasifika cadet journalists with sessions on verification (focused on fake images and information). In addition, YouTube is a supporter of Netsafe's programs and initiatives to reduce online harm. [Updates](#) like these can be found on the [Google New Zealand Blog](#), with the latest efforts featured [here](#).

Google also is a member of the [Digital Trust & Safety Partnership](#), a first-of-its-kind partnership with other leading technology companies committed to developing industry best practices, verified through internal and independent third-party assessments, to ensure consumer trust and safety when using digital services. As discussions on these important issues continue, the group will engage with consumer and user advocates, policymakers, law enforcement, relevant NGOs and various industry-wide experts to help us develop these best practices. The Partnership will share a state-of-the-industry report that evaluates companies' implementation of the practices.

In addition to our publicly available transparency reports (see Outcome 11), YouTube is equipping researchers from around the world with data, tools, and support to advance the public's understanding of our platform and its impact through our [YouTube Researcher Program](#). Eligible researchers from diverse disciplines affiliated with an accredited higher-learning institution can apply to use YouTube data to study a variety of topics. Participants can access the following:

- Scaled access to YouTube's public data corpus with as much quota as required for their



research.

- Opportunity to derive insights from global YouTube data.
- Support and technical guidance from YouTube.

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

As per Measure 45 of the Code, Google is committed to selecting an independent third-party organisation to review its annual compliance reports under the Code and evaluate the level of progress made against the Commitments, Outcomes and Measures, as outlined in section 4 of the Code, as well as commitments made by Signatories in their Participation Form.