

# Aotearoa New Zealand Code of Practice for Online Safety and Harms

## *Google Annual Report*

<b>Signatory:</b>	Google
-------------------	--------

<b><i>If applicable:</i></b> <b>Relevant Products / Services:</b>	YouTube
--	---------

We are pleased to provide our first Annual Report under the Aotearoa New Zealand Code of Practice for Online Safety and Harms. This Report builds on the information provided in our [Baseline Report of 2022](#).

### 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.

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 Baseline Report 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, politics, 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.

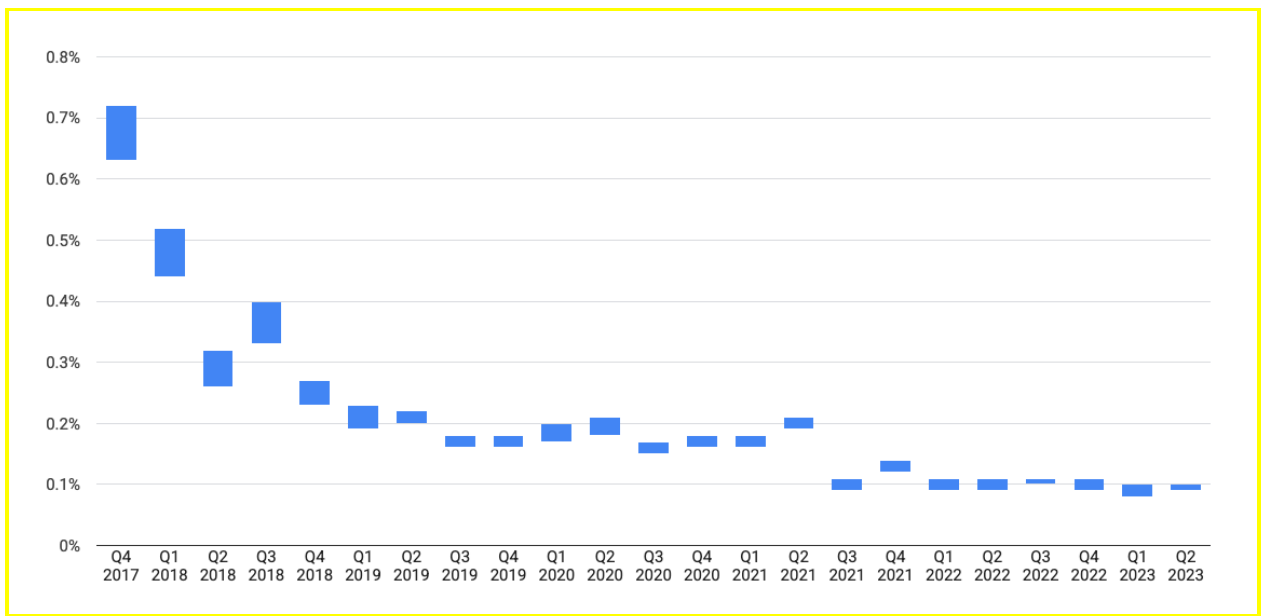
Content that violates our policies will be removed, and repeated violations or a single case of severe abuse may result in channel terminations.

**YouTube Community Guidelines enforcement globally, July 2022 - June 2023**

	Q3 2022	Q4 2022	Q1 2023	Q2 2023
YouTube channels removed globally for violating our Community Guidelines	5,820,978	6,449,291	8,749,977	14,063,789
YouTube videos removed globally for violating our Community Guidelines	5,603,794	5,663,175	6,487,896	7,365,556
YouTube videos removed globally where first detected by automated flagging	5,296,959	5,344,444	6,082,092	6,857,780
Percentage of YouTube videos removed globally before they received more than 10 views	67.9%	68.7%	72.3%	73.4%
Comments removed globally for violating our Community Guidelines	737,512,355	1,932,418,181	853,275,342	767,476,193
Percentage of removed comments globally first detected by automated flagging	99.0%	99.5%	99.4%	99.3%

- 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 2022** (July - Sept 2022), VVR was 0.10 - 0.11% (i.e., out of every 10,000 views on YouTube, 10-11 were of violative content).
  - In **Q4 2022** (Oct - Dec 2022) VVR was 0.09 - 0.11% (i.e., out of every 10,000 views on YouTube, 9-11 were of violative content).
  - In **Q1 2023** (Jan - Mar 2023), VVR was 0.08 - 0.10% (i.e., out of every 10,000 views on YouTube, 8-10 were of violative content).
  - In **Q2 2023** (Apr - Jun 2023), VVR was 0.09 - 0.10% (i.e., out of every 10,000 views on YouTube, 9-10 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 2022 to June 2023, the following actions were taken globally:

	<b>Q3 2022</b>	<b>Q4 2022</b>	<b>Q1 2023</b>	<b>Q2 2023</b>
YouTube channels removed for violating our nudity and sexual content policy	193,730	129,613	163,054	615,725
YouTube channels removed for violating our child safety policy	105,704	92,172	102,357	94,120
YouTube videos removed for violating our nudity and sexual content policy	769,692	682,899	664,643	746,089
YouTube videos removed for violating our child safety policy	2,017,231	1,923,889	2,222,925	2,508,088

### **New Zealand Removals**

In H1 2023 (Jan - Jun 2023), the following actions were taken in respect of YouTube videos uploaded from IP addresses in New Zealand:

<b>4,240</b>	YouTube videos uploaded from IP addresses in New Zealand were removed for violating our child safety policy
<b>725</b>	YouTube videos uploaded from IP addresses in New Zealand were removed for violating our nudity and sexual content policy

## **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 New Zealand's:

In H2 2022 (Jul - Dec 2022)<sup>1</sup>:

<b>359,931</b>	Total pieces of content reported to NCMEC by YouTube
<b>238,827</b>	Cyber Tipline reports to NCMEC by YouTube

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

- YouTube has expanded support for YouTube Supervised Experience so that tweens can sign in to YouTube on gaming consoles, smart TVs, the YouTube Music app and more. This helps us respect a parent’s choice of content setting across devices in the home where their tween watches videos or listens to music.

**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**

YouTube’s [Community Guidelines](#) prohibit content or behaviour intended to harass, threaten or bully others.

- 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 2022 to June 2023, the following actions were taken globally:

	<b>Q3 2022</b>	<b>Q4 2022</b>	<b>Q1 2023</b>	<b>Q2 2023</b>
YouTube channels removed for harassment and cyberbullying	59,884	38,402	36,228	29,940
YouTube videos removed for harassment and cyberbullying	561,185	472,614	494,351	483,144

<sup>1</sup> Data for H1 2023 is not available at the time of submission. Updated data will be published [here](#).

## New Zealand Removals

In H1 2023 (Jan - Jun 2023), the following actions were taken in respect of YouTube videos uploaded from IP addresses in New Zealand:

826

YouTube videos uploaded from IP addresses in New Zealand were removed for harassment and cyberbullying

## Product, Policy and Partnership Updates

- **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 a new, [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 comments that may be considered more hurtful than others 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.
  - Finally, we started rolling out [Channel Guidelines](#), which allows creators to clearly communicate what is and is not ok in their comments section.
- In collaboration with 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 can now access all this information in our new [Creator Safety Center](#).
- **Introduction of comment removal warnings and timeouts:** We have also [introduced](#) comment removal warnings and timeouts. This feature reminds users when we’ve detected and removed some of their comments for violating our Community Guidelines. If a user continues to leave multiple abusive comments, they may receive a timeout and be temporarily unable to comment for up to 24 hours. Our testing has shown that these warnings and timeouts reduce the likelihood of users leaving violative comments again.

**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.

If content directed against an individual is not covered by our hate speech policy, it 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 2022 to June 2023, the following actions were taken globally:

	<b>Q3 2022</b>	<b>Q4 2022</b>	<b>Q1 2023</b>	<b>Q2 2023</b>
YouTube channels removed for being hateful or abusive	35,568	50,822	51,014	21,395
YouTube videos removed for being hateful or abusive	156,676	217,149	177,921	191,080

#### **New Zealand Removals**

In H1 2023 (Jan - Jun 2023), the following actions were taken in respect of YouTube videos uploaded from IP addresses in New Zealand:

**973**

YouTube videos uploaded from IP addresses in New Zealand were removed for being hateful or abusive

**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.
- We've expanded our policies to combat violent extremism by removing content glorifying violent acts for the purpose of inspiring others to commit harm, fundraise, or recruit, even if the creators of such content are not related to a designated terrorist group.
- We do not allow content that includes instructions to kill or harm or which promotes or glorifies violent tragedies under our [harmful or dangerous content policy](#).
- 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. We may make exceptions for content that has a clear educational, documentary, scientific, or artistic purpose. For example, we may allow content depicting terrorist violence published by major news outlets, but if the content is graphic, we may place it behind an interstitial warning to users that the content is graphic in nature (e.g., documentary footage from a war zone). We think this policy is critical to striking the right balance with free expression, and serves important societal purposes. Note that we do not allow the following kinds of content even if there's educational, documentary, scientific, or artistic context provided:

- 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 2022 to June 2023, the following actions were taken globally:

	<b>Q3 2022</b>	<b>Q4 2022</b>	<b>Q1 2023</b>	<b>Q2 2023</b>
YouTube channels removed for the promotion of violence and violent extremism	9,520	10,481	9,492	12,868



YouTube videos removed for the promotion of violence and violent extremism	67,516	74,034	77,038	104,795
--	--------	--------	--------	---------

### New Zealand Removals

In H1 2023 (Jan - June 2023), the following actions were taken in respect of YouTube videos uploaded from IP addresses in New Zealand:

<b>118</b>	YouTube videos uploaded from IP addresses in New Zealand were removed for the promotion of violence and violent extremism
------------	---

### 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 Baseline Report, GIFCT's [Content Incident Protocol](#) has been activated an additional two times in response to the September 2022 attack in Memphis, Tennessee and the April 2023 attack in Louisville, Kentucky. GIFCT has further developed its [Incident Response Framework](#) to include a Content Incident tier to respond to non-live-streamed, perpetrator-produced video and images depicting the real-world event. This was activated for the first time following an attack in [Udaipur](#), Rajasthan, India (July 2022).

### 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 content depicting terrorist violence published by major news outlets but if the content is graphic, we may place it behind an interstitial warning to users that the content is graphic in nature (e.g., documentary footage from a war zone). We think this policy is critical to striking the right balance with free expression, and serves important societal purposes. Note that we do not allow the following kinds of content even if there's educational, documentary, scientific, or artistic context provided:

- 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 2022 to June 2023, the following actions were taken globally:

	<b>Q3 2022</b>	<b>Q4 2022</b>	<b>Q1 2023</b>	<b>Q2 2023</b>
YouTube channels removed for violent or graphic content	51,051	36,230	19,450	14,685
YouTube videos removed for violent or graphic content	1,032,392	930,472	1,011,206	946,182

**New Zealand Removals**

In H1 2023 (Jan - June 2023), the following actions were taken in respect of YouTube videos uploaded from IP addresses in New Zealand:

**564** YouTube videos uploaded from IP addresses in New Zealand were removed for violent or graphic content

**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 manipulated and misattributed content that may pose a serious risk of egregious harm, and misleading content that suppresses census participation.
- [Election 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.
- [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.

## Enforcement

### Global Removals

From July 2022 to June 2023, the following actions were taken globally:

	Q3 2022	Q4 2022	Q1 2023	Q2 2023
YouTube channels removed for misinformation <sup>2</sup>				92,976
YouTube videos removed for misinformation	121,423	100,016	69,929	78,413

### New Zealand Removals

In H1 2023 (Jan - Jun 2023), the following actions were taken in respect of YouTube videos uploaded from IP addresses in New Zealand:

<b>338</b>	YouTube videos uploaded from IP addresses in New Zealand were removed for misinformation
------------	--

<sup>2</sup> Beginning with the April-June 2023 reporting period, we began to break out data for channels removed under our policies related to “Misinformation”, including our medical misinformation policies, certain types of technically manipulated content, and content interfering with democratic processes. (Note: These are just examples; other content may now or in future be classified under this category.) In previous reporting periods, these types of removals were nested under the “Spam, misleading and scams” category. We also now nest the previous “Impersonation” category for channel removals under the category “Misinformation”.

## ***Product, Policy and Partnership Updates***

- **Launch of information panels in New Zealand:** In March 2023, we [announced](#) that we were expanding [information panels giving topical content](#) to New Zealand. Now, when people search videos on certain topics prone to misinformation (such as climate change), they'll see an information panel at the top of their search results or under a video they're watching which includes links to additional info and context from authoritative third-party sources.
- **Updated approach to medical misinformation:** YouTube recently streamlined dozens of our existing medical misinformation guidelines to fall under three categories – Prevention, Treatment, and Denial. These policies apply to specific health conditions, treatments, and substances where content contradicts local health authorities or the WHO. To determine if a condition, treatment or substance is in scope of our medical misinformation policies, we'll evaluate whether it's associated with a high public health risk, publicly available guidance from health authorities around the world, and whether it's generally prone to misinformation. Further details on the framework can be found [here](#).
- **Launch of Hit Pause, YouTube's global media literacy campaign:** In 2022, YouTube launched [Hit Pause](#), an educational program seeking to teach viewers critical media literacy skills.
- **Training New Zealand journalists to spot misinformation ahead of the New Zealand General Election:** In recent years, Google has trained 935 journalists across New Zealand on a range of digital tools, and we provide a comprehensive [online training website](#). As part of our ongoing commitment to the New Zealand media industry Google has partnered with CrossCheck at RMIT FactLab to launch [RMIT Fact Lab Workshops](#). These immersive and interactive trainings help newsrooms and community media learn skills to analyse online information during the Election period. See [here](#) for more information.
- **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. More information can be found [here](#).
- **Access to information during an Election Period:** In the build-up to elections, people need helpful and authoritative information to help them navigate the electoral process. For the New Zealand election in 2023, we are making electoral information from the [Election Commission](#) and other authoritative sources - such as how to vote and where to vote - easily discoverable on YouTube. 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 Squiz Kids'](#) media literacy module, "Newshounds" to launch its plug and play resources for teachers, children and their parents in New Zealand.
  - Partnered with Te Rito to continue our efforts to fund and provide a [training camp for 30 Māori cadet journalists](#) with sessions on verification (focused on fake images and information).

- Partnered with Telum Media to run [Digital Skills Training for New Zealand journalists](#), using Google’s strengths and insights to train journalists in how to find and present stories to engage audiences.
- Sponsored an [online tracking tool](#) for journalists to understand, monitor and report on ahead of New Zealand’s local elections.
- Sponsored [Trusted Media Summit \(APAC\)](#), where news partners build a community to understand the trends in misinformation and audience behaviour.

**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 2022 to June 2023, the following actions were taken globally:

	Q3 2022	Q4 2022	Q1 2023	Q2 2023
YouTube channels removed for spam, misleading and scams	5,305,836	6,022,413	8,293,775	13,147,984
YouTube channels removed for impersonation <sup>3</sup>	31,279	36,710	34,725	

<sup>3</sup> See footnote 2 above.

YouTube videos removed for spam, misleading and scams	218,885	260,864	342,567	963,983
---	---------	---------	---------	---------

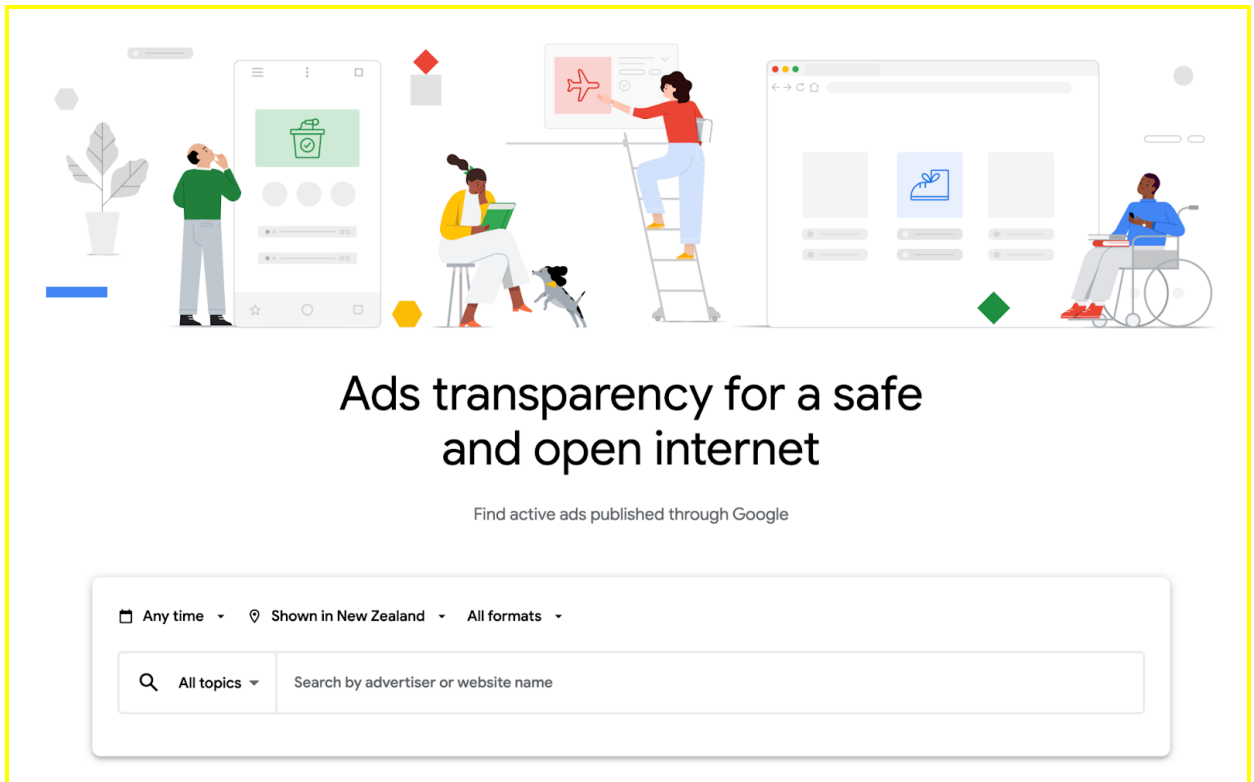
### New Zealand Removals

In H1 2023 (Jan - Jun 2023), the following actions were taken in respect of YouTube videos uploaded from IP addresses in New Zealand:

<b>774</b>	YouTube videos uploaded from IP addresses in New Zealand were removed for spam, misleading and scams
------------	--

### Product, Policy and Partnership Updates

- **Launch of the Ads transparency Centre:** On 29 March 2023, Google announced the [Ads Transparency Centre](#) to help users quickly and easily learn more about the ads you see from Google, including Search, YouTube and Display. The Ads Transparency Center is a searchable hub of all ads served from verified advertisers. This one-stop shop was designed with you in mind by ensuring you have easy access to information about the ads you see from Google. With the Ads Transparency Center, you'll be able to understand:
  - The ads an advertiser has run
  - Which ads were shown in a certain region
  - The last date an ad ran, and the format of the ad.



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 are disclosed in the Quarterly Bulletin (published on our [Threat Analysis Group blog](#)). Examples of action taken against coordinated influence operation campaigns in Q2 2023 include:

  - In April, we terminated 139 YouTube channels as part of our investigation into coordinated influence operations linked to Ukraine. The campaign was sharing content in Russian that was supportive of Russia and Kazakhstan President Kassym-Jomart Tokayev.
  - In May, we terminated 337 YouTube channels as part of our investigation into coordinated influence operations linked to Russia. The campaign was linked to a Russian consulting firm and was sharing content in Russian that was supportive of Russia and Russian President Vladimir Putin and critical of Ukraine, NATO and the West.
  - In June, we terminated 6,575 YouTube channels as part of our ongoing investigation into coordinated influence operations linked to China. These channels mostly uploaded spammy content in Chinese about music, entertainment and lifestyle. A very small subset uploaded content in Chinese and English about China and U.S. foreign affairs. These findings are consistent with our previous reports.

Further information can be found in the [TAG Bulletin: Q2 2023](#).

## 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

With over 500 hours of content uploaded to YouTube every minute, 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 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.

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 Baseline Report, including:

- [YouTube Kids](#) and [YouTube supervised experience](#) for families and younger users.
- [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.
- [Restricted mode](#) is an optional setting created to give viewers better control over the content that they see by helping to screen out potentially mature content that users may prefer not to view.



**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, using restricted mode 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.

YouTube also 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.

### **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 Baseline Report 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.

## Updates

- **Blogpost on policy development at YouTube:** In December 2022, we published a blogpost answering commonly asked questions about how we develop our Community Guidelines. The blogpost covers issues such as how we determine what policy updates are needed, how we decide where to draw the line, who provides input on policy development and enforcement, whether we try to get ahead of emerging issues and how we make sure our policies are enforced consistently. The blogpost can be found [here](#).
- **Launch of the Transparency Centre:** In August 2023, we announced the [Transparency Centre](#), 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](#).

**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 Baseline Report provides 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.
- 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](#).

### Updates

- **As noted under Outcome 10 above**, the [Transparency Centre](#) is a central hub providing easy access to information on our policies, how we create and enforce them, including our transparency reports.

## 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" is a plug and play resources for teachers, children and their parents in NZ; and with Te Rito to fund and provide a training camp for Māori 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.

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.