



Google Australia Pty Ltd
48 Pirrama Road
Pyrmont, NSW 2009

(02) 9374 4000 main
Google.com

Questions tendered at the Standing Committee on Social Media and Online Safety

Question 1

Your submission says you have “strict policies and robust operations in place to tackle content and behaviour that is harmful or exploitative to children”. The Committee has heard evidence that paedophiles were able to network in the comments sections of children on Youtube and were able to post time stamps of moments in these videos that they found arousing to highlight for other paedophiles. This behaviour led to Youtube’s recommendation engine recommending other videos of children to these child abusers. Ultimately scores of videos recorded tens of millions of views before this abusive use was discovered by child abuse campaigners.

- Why did Youtube not identify this behaviour itself?
- Why were there not safeguards on Youtube’s automated recommendation engine to stop recommendations relating to videos featuring children?

Response

Since 2019 when the events in this report occurred, we have disabled comments on hundreds of millions of videos featuring minors across the platform, to limit the risk of exploitation. Additionally, we implemented a classifier that scans comments for child safety violations and auto-removes hundreds of thousands a day.

Over the last few years, we’ve made regular improvements to the machine learning classifier that helps us protect minors and families. With this classifier, we’re able to better identify videos that may put minors at risk and apply our protections, including disabling comments and limiting recommendations on hundreds of millions of videos, across our platform.

Last year, building on our industry-leading transparency initiatives, Google launched a transparency report¹ specifically dedicated to detailing our efforts to combat online child sexual abuse material. From January through June 2021, these efforts included the submission

¹ <https://transparencyreport.google.com/child-sexual-abuse-material/reporting?hl=en>

of over 400,000 CyberTipline reports to NCMEC, involving over 3.4 million pieces of content. YouTube in particular contributed more than 124,000 CyberTipline reports and over 133,000 pieces of content to NCMEC.

When it comes to kids, we take an extra cautious approach towards our enforcement. Our Community Guidelines expressly prohibit “sexually explicit content featuring minors and content that sexually exploits minors.” We also remind uploaders, “before posting videos of yourself, your family, or friends, think carefully about whether it may put anyone at risk of negative attention.”

In addition to our long-standing efforts to combat Child Sexual Abuse Material (CSAM) videos, we have made large investments to detect and remove content which may not meet the legal definition of CSAM, but where minors are still being sexualised or exploited. We have policies against videos, playlists, thumbnails and comments that sexualise or exploit children (more information is available at <https://support.google.com/youtube/answer/2801999>).

We use machine learning systems to proactively detect violations of these policies and have human reviewers around the world who quickly remove violations detected by our systems or flagged by users and our trusted flaggers. Our machine learning systems help to proactively identify videos that may put minors at risk and apply our protections at scale, such as restricting live features, disabling comments, and limiting video recommendations. Identifying and removing videos more quickly—often before they have even been viewed—means children who are being sexually abused today are more likely to be identified and protected from further abuse.

Additionally, we don’t allow users under 13 to create accounts on YouTube.com unless they are [Supervised Accounts](#) set up by a parent / carer. In cases where we identify an account that may be owned by someone who may be under 13, we terminate it. We terminate tens of thousands of accounts every week as part of this process.

Finally, we work with the industry by offering expertise and technology (e.g. CSAI Match) to smaller partners and NGOs. We encourage user flags and invite specialist NGOs to flag content to us via our Trusted Flagger program. This program includes a number of child safety focused NGOs.

Question 2

Your submission identifies nine separate regulatory processes in addition to this inquiry currently being undertaken by the Morrison government in this space. You’ve pointed out that five of these processes are proposing the implementation of age verification requirements on platforms. In response, you’ve called for a ‘coordinated approach to these issues that is unified

under a whole of government approach'. What are the risks to good policy outcomes of this government operating parallel, overlapping policy development processes?

Response

An unintended consequence of having such a diversity of approaches is that it risks duplicative and overlapping obligations on industry or inconsistent obligations across different regulations, making it potentially implausible and unreasonably difficult and burdensome to transform theory into practice. It also creates regulatory uncertainty for all businesses impacted by the regulations and risks deterring them from either establishing themselves, or further investing, in Australia. This would jeopardise the Government's stated intention to be a leading digital economy by 2030.

Question 3 (from Tim Watts)

What are Google's views on the recent FTC proposal for oversight on algorithms?

Response

While it is unclear what specific proposal is being referenced, researchers can currently use YouTube's Application Programming Interface (API), and we are actively collecting feedback to better support researcher access to data. Researchers have used that API to publish numerous studies about how YouTube works and its algorithms and systems.

We've invested significantly in other forms of transparency – for instance in user controls over the inputs and outputs of recommendations, and in publicly available resources online. For example, [How YouTube Works](#) explains our approach to recommendations, user settings, and more:

- General information on signals: We're constantly testing, learning and adjusting to recommend videos that are relevant to our users. We take into account many signals, including your watch and search history (if enabled) as well as the channels that you've subscribed to. We also consider your context, such as your country and time of day. For example, this helps us show you locally relevant news. Another factor that YouTube's recommendation systems consider is whether others who clicked on the same video watched it to completion – a sign that the video is higher quality or enjoyable – or just clicked on it and shortly after starting to view the video, clicked away.*
- User feedback: We also ask users directly about their experience with individual videos using random surveys that appear on their homepage and elsewhere throughout the app. We use this direct feedback to fine-tune and improve our recommendations for all users.*

- *User controls: We also provide ways for you to tell us when we're recommending something that you aren't interested in. For example, buttons on the homepage and in the 'Up next' section allow you to filter and choose recommendations by specific topics. You can also click on 'not interested' to signal to YouTube that a video or channel is not what you wanted to see at that time. We provide links to [manage your recommendations and search results](#).*
- *Reducing harmful content in recommendations: We provide insights to users, e.g. via [blog posts](#), on our reducing recommendations of borderline content and content that could misinform users in harmful ways—such as videos claiming the earth is flat, or making blatantly false claims about historic events like the Holocaust.*

Question 4 (from Craig Kelly)

Does Alphabet invest in / have financial holdings in any COVID vaccination companies?

Response

[Google Ventures](#) (GV) is an independent, return-driven fund. Today GV has more than \$5 billion under management, with Alphabet as its sole limited partner. GV invests across all stages and sectors, with a focus on enterprise, life sciences, consumer, and frontier technology. A list of those investments, including in Vaccitech, are publicly available and can be found at the following link: <https://www.gv.com/portfolio>

Question 5 (from Celia Hammond)

You've mentioned that there are 2,000 staff in Australia. I would like the proportion of that in equivalent full-time, EFT, terms, as opposed to just bodies; how many of them work in the safety team; and the qualifications and criteria behind them. You've also mentioned that there are 20,000 people worldwide working in your safety team. I'd like that information in terms of equivalent full time and, as a comparison and a metarate, against how many work in advertising, marketing and systems development—they can be globalised—as a percentage of your entire workforce, again with their training and the inherent requirements of their job.)

Response

We employ around 2,000 full time permanent employees in Australia. Last year, Google had approximately 20,000 employees working globally in a variety of different roles to help enforce our policies and moderate content.

Question 5 (ct'd; from Celia Hammond)

How many of the 20,000 people working on content moderation are full time employees?

Response (ct'd)

We employ teams all around the world to work on various aspects of content safety. The exact allocation of the work to full time employees varies over time -- for instance, crises or new threats can require some teams to drop existing work or others to ramp up for rapid response.

Question 5 (ct'd; from Celia Hammond)

What percentage of Google's entire global workforce work on advertising / sales?

Response (ct'd)

Personnel sitting in various teams may conduct work that relates to Google's advertising and sales efforts, making it difficult to provide an accurate estimate of the percentage of Google's entire global workforce that works on advertising / sales. Google's Global Business Organisation is the team that is primarily focused on our sales and advertising business, and as of December 2021, it accounted for approximately 16% of Google's workforce.

Question 5 (ct'd; from Celia Hammond)

Do Google's executive contracts include KPIs that relate to online safety / security?

Response (ct'd)

Google has a longstanding commitment to online safety / security and this is consistently expressed in our [founders letters](#) and in our company annual reports. Additionally, Alphabet's CEO, Sundar Pichai, has [testified](#) that "keeping users safe and secure on our platforms is a top priority," and YouTube CEO, Susan Wojcicki identified protecting the YouTube community as one of [YouTube's top priorities for 2022](#).

Question 6 (from Celia Hammond)

What proactive schemes does Google run to conduct reviews of its online safety / security processes and policies? Are these schemes audited (either internally or externally)? If so, how often and what does Google spend on these audits?

Response

Google has long been committed to and prioritised online safety and security, as expressed in our [founders letters](#) and in our company annual reports (see question 5). We invest on an ongoing basis in quality testing for our services and in assessments of our policies and enforcement so as to spot issues or unearth vulnerabilities that may result from new threats and changes to the online content ecosystem. This is a key part of the effort of the more than 20,000 employees across Google products and services, including YouTube, whose work relates to content moderation.

Question 7 (from Lucy Wicks)

How many FTEs in Australia are wholly focused on women / children's online safety and what percentage of our total staff numbers?

Response

There are a number of employees who work on online safety for women and children (and online safety generally) for Australia as part of our team of experts on these issues.

We also have many horizontal teams that contribute to the broader infrastructure needed to promote safety online. This infrastructure includes, but is not limited to, our heavy investment in artificial Intelligence and machine learning technologies that are used across many of Google's products and services.

Question 8 (from Lucy Wicks)

[In a survey,] 30 per cent of women who were surveyed indicated they'd experienced online abuse or harassment, and 42 per cent of women who had experienced online abuse said it was sexist or misogynist in nature and 20 per cent said it included threats of physical or sexual violence. Could you provide your specific and detailed approach to addressing those concerns with individuals, community groups and public figures?

Response

The YouTube Community Guidelines prohibit harassment and cyberbullying, hate speech. YouTube's hate speech policy outlines clear guidelines prohibiting content that promotes violence or hatred against individuals or groups based on certain attributes, such as gender identity and expression. Harassment and cyberbullying are not allowed on YouTube, and we have clear policies that prohibit content targeting an individual with prolonged or malicious insults or threats based on certain attributes, such as gender identity and expression.

Individuals / organisations can report instances of potentially violative content directly to us using the abuse reporting tools that we have built into all Google products (e.g. flagging videos on YouTube).

We encourage users to flag individual videos for our attention, or contact us about inappropriate comments using the Help & Safety Tool. YouTube also has an online webform for users to file privacy or defamation complaints, which are evaluated on a case-by-case basis. We also give people tools for blocking other users, preventing them from making comments on your videos or from contacting you through private messages.

We regularly review our policies and products to see how we can better protect the YouTube community. We communicate with third-party groups, such as the Women's Services Network (WESNET), and incorporate their feedback accordingly. We also provide creators with control over their comments settings and enable them to block commenters.

If user content violates our policies, we'll remove the content and notify the user via email. We may terminate a channel or account for repeated violations of the Community Guidelines or Terms of Service, per our three-strikes policy. We may also terminate a channel or account after a single case of severe abuse, or when the channel is dedicated to a policy violation. Additionally, we may remove content or issue other penalties—such as terminating an account—when a creator repeatedly targets, insults and abuses a group based on attributes such as race, ethnicity, sexual orientation, or gender identity and expression, across multiple uploads.

Question 9 (from Lucy Wicks)

When asked at recent public hearings whether certain statements would breach a platform's terms of service and allow for removal, a number of social media companies outlined that the answer depended on the context of the comments. Can you outline to the committee the context in which these statements would not breach your policies and remain on your site? Would the context differ for a private individual and a public figure?

Response

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

Our policies set out what's allowed and not allowed on YouTube, and apply to all types of content on our platform, including videos, comments, links, and thumbnails. For example, our

harassment and cyberbullying [policies](#) prohibit content that threatens individuals or that targets an individual with prolonged or malicious insults based on intrinsic attributes. Our hate speech [policy](#) prohibits violence or hatred against individuals or groups based on attributes including age, ethnicity, race, disability, gender, religion, amongst others.

Our Community Guidelines are intended to balance the importance of allowing or providing access to a broad range of diverse opinions and perspectives, particularly in the context of political or public debate, with preventing harm that may be caused by such content. This may mean that relevant content is left up even where some may find it offensive, controversial or disagreeable.

The context matters, as there may be offensive content which is intended for educational, documentary, scientific or artistic purposes. For example, as an exception to our policies on graphic violence, we may allow footage taken by a citizen in a warzone, recognising there may be a documentary public interest in that content. Or a satirical video may use offensive language in the context of showing real world events and incidents. We can and do apply a warning and age restriction on certain content even where these exceptions are applied. These decisions are nuanced, context is important, and we take care to make these decisions by looking at multiple factors, including the context in which the video is posted, the video title, descriptions and other context.