Google Australia Pty Ltd
Level 5, 48 Pirrama Road
Pyrmont, NSW 2009
Australia

google.com

Committee Secretary
Parliamentary Joint Committee on Intelligence and Security
PO Box 6021
Parliament House
Canberra ACT 2600

BY EMAIL: pjcis@aph.gov.au

Friday 12 March 2021

INQUIRY INTO EXTREMIST MOVEMENTS AND RADICALISM IN AUSTRALIA

Google welcomes the Committee's inquiry into extremist movements and radicalism within Australia. The use of digital platforms by extremist groups to communicate, congregate and promote their ideology is of grave concern to Google. Tackling this issue requires a whole of society effort across Government, law enforcement, community organisations, mental health support services, schools and the technology industry. As a technology company, we take these issues seriously and we want to be a part of the solution.

We believe the free flow of information and ideas has important social, cultural and economic benefits, though society has always recognised that free speech must be subject to reasonable limits. This is true both online and off, and it is why, in addition to respecting the law, we have additional policies, procedures, and community guidelines that identify what activity is not permissible on our platforms. In recognition of the fact that addressing these issues is a shared responsibility, we work closely with government, industry, and civil society to address these challenges.

YouTube and Google hosted products have policies against violent extremism and we prohibit designated terrorist groups from posting any content. Our efforts in this space have focused on five key areas:

1. Detection and removal;
2. Working with experts;
3. Amplification and promotion of counter-messaging;
4. Collaboration with law enforcement agencies; and
5. Working within multilateral fora to share information and best practices.

Improving detection and ensuring faster removal (powered by machine learning)

We have robust policies and programs to defend our platforms against the spread of hate and any incitement to violence. This includes prohibitions on: terrorist recruitment, violent extremism, incitement to violence, glorification of violence, and instructional videos related to acts of violence. We apply these policies to violent extremism of all kinds, whether inciting violence on the basis of race or religion or as part of an organised terrorist group.

In order to improve the effectiveness of our policy enforcement, we have invested heavily in both technology and people to quickly identify and remove content that violates our policies against incitement to violence and hate speech:

1. YouTube's enforcement system starts from the point at which a user uploads a video. If our technology detects that the video is similar to videos that we know already violate our policies, it is sent for humans to review. If they determine that it violates our policies, they remove it and the system makes a "digital fingerprint" or hash of the video so it can't be uploaded again.
2. Machine learning technology also helps us more effectively identify this content and enforce our policies at scale. We introduced machine learning technology to detect extremist content in June 2017. To train our machine learning classifiers, our teams hand-reviewed over 2 million pieces of content. In turn, we calculated that from June-Dec. 2017, the technology reviewed and flagged content that would have taken 180,000 people working 40 hours a week to assess. This technology continues to learn and improve over relatively short periods of time.
3. This broad cross-sectional work has led to tangible results that we publish quarterly in the YouTube Community Guidelines Enforcement Report. Over 94% of the 9.32 million videos we removed in the Oct-Dec quarter of 2020 were first flagged by our automated systems[1]. Almost 72% of these videos were removed before they reached ten views. And overall, videos that violate our policies generate a fraction of a percent of the views on YouTube.

In terms of human resources, we employ approximately 20,000 people across Google working to tackle abuse, including engineers, reviewers, and subject-matter experts. Our review teams are located in countries around the world, are fluent in multiple languages, and carefully evaluate flags 24 hours a day, seven days a week.

Lastly, YouTube's search and recommendation systems reflect what people search for, the number of videos available, and the videos people choose to watch on YouTube. We pull in recommendations from a wider set of topics--on any given day, more than 200 million videos get recommended on the YouTube homepage alone. Diversity of information is built into the design of YouTube -- each query delivers multiple options from various sources. Users consume a diverse array of content. Recommendations are designed to deliver new content to users. We use various signals to suggest videos via watch next. We've also begun recommending videos from more authoritative sources in certain content verticals, as well as

[1] https://transparencyreport google com/youtube-policy/removals?hl=en

providing information panels to provide users with more information and context when they search for or watch certain content.

In the last couple of years, we've made hundreds of individual changes to improve the quality of recommendations on YouTube. In 2019, we started taking a tougher stance on videos that contain inflammatory or supremacist content, leading to more takedowns, demonetisation, and ineligibility to surface in recommendations. Where content clearly crosses the line into hate speech, we remove it, and we terminate the channels of repeat violators.

Working with expert partners to help understand the issues and identify violative content

As hate and violent extremism content is constantly evolving and can sometimes be context-dependent, we also rely on experts to help us identify policy-violating videos. Some of these experts sit on our internal 'intel desk', which proactively looks for new trends in content that might violate our policies.

Outside the company, we have expanded the network of experts that we work with to gain a better understanding of emerging trends, patterns of behaviour and intervention strategies. This network includes academics, NGOs, civil society groups and of course Governments; some of whom participate in the YouTube Trusted Flagger program. Trusted Flaggers are given bulk flagging tools, and their flags are actioned for priority review. We now have a global network of over 200 Trusted Flaggers drawn from NGOs and government agencies, with several that specialise in violent extremism, including The International Centre for the Study of Radicalisation and Political Violence, the Institute for Strategic Dialogue, the Wahid Institute in Indonesia, and local government agencies focused on counterterrorism, including the Australian Federal Police, the Department of Home Affairs and the Office of eSafety Commissioner. We benefit immensely from their expertise.

Alphabet's Jigsaw group[2], an incubator to tackle some of the toughest global security challenges, has deployed the Redirect Method, which uses targeting tools and curated YouTube playlists to disrupt online radicalisation. The method is open to anyone to use, and NGOs have sponsored campaigns against a wide-spectrum of ideologically-motivated terrorists and violent extremists. We ran a pilot of the Redirect Method here in Australia back in 2017 in partnership with All Together Now and People Against Violent Extremism (PAVE).

Counter messaging

We have also invested in programs and projects to support counter-messaging campaigns and community initiatives that promote tolerance and understanding. These programs present narratives and elevate credible voices speaking out against hate, violence, and terrorism. For example, the YouTube Creators for Change program[3] supports creators who are tackling tough issues, including extremism and hate by building empathy and acting as positive role models. To complement the global Creators for Change campaigns, we launched

[2] https://jigsaw.google.com/

[3] https://www.youtube.com/creators-for-change/

several Australia specific programs between 2017-2020[4] resulting in the creation of almost 30 unique pieces of content by Australian content creators.

Google has donated millions of dollars to nonprofits around the world[5] seeking to empower and amplify counter-extremist voices.

Collaboration with law enforcement agencies

Google appreciates that law enforcement agencies face significant challenges in protecting the public against crime and terrorism. Google engages in ongoing dialogue with law enforcement agencies to understand the threat landscape and respond to threats that affect the safety of our users and the broader public. We have a well established process in place for law enforcement agencies to lawfully request user data from Google and we receive approximately 4,000 requests each year from Australian agencies. When we become aware of statements on our platform that constitute a threat to life or that reflect that someone's life may be in danger, we report this activity to law enforcement agencies.

Under U.S. law, the Stored Communications Act allows Google and other service providers to voluntarily disclose user data to governmental entities in emergency circumstances where the provider has a good faith belief that disclosing the information will prevent loss of life or serious physical injury to a person. Our team is staffed on a 24/7/365 basis to respond to these emergency disclosure requests (EDRs). We have seen significant growth in the volume of EDRs that we receive from governmental entities, as illustrated in our transparency report covering government requests for user data[6].

Encryption is a critically important tool in protecting users from a broad range of threats. We design our products and services with advanced security at their core -- from our custom-built infrastructure that protects our data centers and servers to layers of advanced encryption that protect user data.

We appreciate that law enforcement and intelligence agencies face significant challenges in protecting the public against crime and terrorism. Our work to increase the cybersecurity posture of users while enabling law enforcement agencies across the world to investigate and solve crimes demonstrates that the goals of public safety and user security are compatible. There are a lot of opportunities to work together to reduce the challenges confronting law enforcement regarding obtaining digital evidence. The CSIS report[7] on "low hanging fruit" remains a good and relevant resource for recommendations to improve the ability of law enforcement to obtain digital evidence.

Strong encryption doesn't create a law free zone -- companies can still deploy several anti-abuse protections using metadata, behavioural data, and new detection technologies --

[4] http://sharesomegood.org/

[5] https://www.blog.google/outreach-initiatives/google-org/supporting-new-ideas-fight-against-hate/

[6] Australian requests for user data available at https://transparencyreport.google.com/user-data/overview?hl=en&user_requests_report_period=series:requests,accounts;authority:AU;time:&lu=user_requests_report_period

[7] https://www.csis.org/analysis/low-hanging-fruit-evidence-based-solutions-digital-evidence-challenge

without seeing the content of messages encrypted in transit (thereby respecting user privacy).

While we are unable to provide to law enforcement the unencrypted content of messages encrypted in transit, we are still able to provide a wealth of data and signals that in some instances have proven richer than content data. Metadata such as call location, associated phone numbers, frequency and length of call/text are logged on our servers and can be shared with law enforcement/intelligence when provided with a valid court order.

Multilateral engagement and cooperation

We are also deeply committed to working with governments, the tech industry, and experts from civil society and academia to protect our services from being exploited by bad actors. The tragic events in Christchurch presented unique challenges, and we had to take unprecedented steps to address the sheer volume of new videos related to the events. In the months following, Google and YouTube signed the Christchurch Call to Action, a series of commitments to quickly and responsibly address terrorist content online.

This is an extension of our ongoing commitment to working with our colleagues in the industry to address the challenges of terrorism online. Since 2017, we've done this through the Global Internet Forum to Counter Terrorism (GIFCT), of which YouTube is a founding company and was its first chair. The GIFCT was designed to foster technical collaboration among member companies (largely through the hash sharing consortium whereby a database of unique digital fingerprinted images and videos of known violent terrorist imagery or recruitment videos or are shared amongst consortium members) , advance relevant research, and share knowledge with smaller platforms who typically have less resources than the larger founding members.

On July 24, 2019, as part of our steps to implement the Christchurch Call, GIFCT announced the introduction of a joint content incident protocol to enable and empower companies to more quickly and effectively respond to emerging and active events. A key component of this protocol is the ability to share content hashes related to an incident so that all GIFCT members can quickly detect and remove the same content if it appears on their platforms. The deployment of the hash sharing database in times of crisis has made an appreciable impact on the virality and availability of terrorist / violent extremist content on mainstream, moderated social media services[8]. The GIFCT also released its first-ever Transparency Report and a new counterspeech campaign toolkit that will help activists and civil society organisations challenge the voices of extremism online.

On 23 September, 2019, the founding GIFCT companies, YouTube, Twitter, Microsoft and Facebook announced that the GIFCT would become an independent organisation with the refreshed mission of preventing terrorists and violent extremists from exploiting digital platforms. On 23 June, 2020 the GIFCT announced Nicholas J. Rasmussen, former Director of the National Counterterrorism Centre, as the first full-time executive director of the organisation.

8 https://www.techagainstterrorism.org/2019/10/15/analysis-what-can-we-learn-from-the-online-response-to-the-halle-terrorist-attack/

Terrorist organisation lists

We define violent extremist content as content made by, or in support of violent criminal organisations; that promotes terrorist acts, including recruitment; or that celebrates or glorifies terrorist attacks. This can include content created by individuals, gangs, known/named terrorist organisations or other groups. Content that violates our policies against violent extremism also includes material produced by government-listed foreign terrorist organisations. We do not permit these terrorist organisations to use our platforms for any purpose, including recruitment.

We are guided by lists of designated terrorist groups compiled by democratically-elected governments with a transparent process for adding and removing groups. We believe democratically elected governments are better positioned to adjudicate the sensitive issue of defining who is or who is not a terrorist. We recognise, however, that there are far fewer government-listed organisations for right-wing terrorism or violent extremism. Therefore activity or content created by individuals or groups that do not appear on any designated terrorist lists will still be assessed according to local law and our terms of service for the product in question. In other words, content that celebrates or promotes terrorist acts, depicts graphic violence or which violates our hate speech policies will continue to be removed regardless of whether the content was created by an organisation that has been designated a terrorist group.

Conclusion

We take the safety of our users very seriously and value our close and collaborative relationships with the broader community of NGOs, law enforcement and government agencies. We have invested substantial resources to tackle the problem of hate speech. At present, we spend hundreds of millions of dollars annually and have more than 20,000 people working in a variety of roles across Google to help enforce our policies and moderate content.

We understand these are difficult issues of great interest to the Australian Parliament and want to be responsible actors who are a part of the solution. As these issues evolve, Google will
continue to invest in the people and technology to meet the challenge. We hope that this submission is of assistance to the Committee as it examines these issues.

Yours sincerely,

[redacted]

Samantha Yorke
Government Affairs and Public Policy