Google Australia Pty Ltd
Level 5, 48 Pirrama Road
Pyrmont NSW 2009
Tel: 02 9374-4000
Fax: 02 9374-4001
www.google.com.au

15 October 2021

Parliamentary Joint Committee on Law Enforcement
PO Box 6100
Parliament House
Canberra ACT 2600
BY EMAIL le.committee@aph.gov.au

Inquiry into the *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019*

Dear Committee members,

Thank you for the opportunity to contribute to this statutory review of the Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019 (the Act). Google has not received any notices under the Act, however we recognise and accept the responsibility that comes with operating a global video sharing platform like YouTube, where 500 hours of new content is being uploaded every minute of the day. Google spent over $1.5 billion dollars on content moderation efforts in 2020 and has nearly twenty-two thousand employees dedicated to ensuring the protection of our platforms. It's a complex task, and–just as in offline contexts—it's not a problem that can be totally solved. Rather, it's a problem that must be managed, and we are constantly refining our policies and processes.

General comments

We wholeheartedly share the Government's intent in seeking to ensure that terrorist and violent extremist content is quickly detected and removed from digital platforms. YouTube and Google hosted products have policies against violent extremism and we prohibit designated terrorist groups from posting any content. YouTube is a signatory to the global Christchurch Call to Action, a series of commitments to quickly and responsibly address terrorist content

online and a founding member of the Global Internet Forum to Counter Terrorism (the GIFCT), the independent organisation delivering on many of the Christchurch Call commitments.

Our efforts to combat illegal content, as well as videos that violate YouTube's Community Guidelines, are reflected in the quarterly published YouTube Community Guidelines Enforcement Report. The most recent report covers April - June 2021 and during this time 6.9% of the 6,278,771 videos removed globally were removed for violating the prohibition on sharing content that promotes violence or violent extremism. Of the 6,278,771 videos removed, 5,927,201 videos were first identified by proactive algorithmic scanning and, of these, 74.6% of videos were detected and removed with less than 10 views. These data points reflect YouTube's ongoing commitment to use technology and people to prevent abhorrent violent content from being distributed at scale across the platform.

We also work closely with companies across the sector through the GIFCT. The GIFCT brings together industry, governments and civil society to foster collaboration and information-sharing to counter terrorist and violent extremist activity online. Through the GIFCT, industry has developed the Content Incident Protocol, a process by which GIFCT member companies quickly become aware of, assess, and address potential content circulating online resulting from an offline terrorist or violent extremist event. Other areas of collaboration include the hash sharing database where members can share hashes of known violent extremist and terrorist content, investment in research and knowledge sharing with smaller platforms who typically have less resources than the larger founding members.

We hope that this review into the Act will examine the entire ecosystem -- recognising that addressing terrorist and violent extremist content is a shared responsibility, and all parties should take proportional, tailored steps to address illegal content. We note that the eSafety Commissioner has issued at least 23 notices under the Act and that 93% of these notices have resulted in the removal of content. We welcome more transparency about the recipients of these notices and specifically those organisations that are not complying with orders issued under the Act, as this will assist the Committee's investigations into the effectiveness of the Act.

At the same time, it is also worth highlighting insights offered by Tech Against Terrorism, an initiative launched and supported by the United Nations Counter Terrorism Executive Directorate. They acknowledge the efforts that have been made to coordinate mainstream tech sector response to content incidents, and highlight that more work is needed to coordinate behaviour across mainstream media, academia, government, and the broader tech industry. In the context of the synagogue attack in Halle in 2019, Tech Against Terrorism observed that both smaller and larger platforms were prompt in dealing with the proliferation of the livestreamed video of the attack and that it was only circulated widely with little or no moderation on smaller fringe platforms. They concluded that responses by smaller and larger tech platforms to prevent virality are being undermined by both the mainstream media and

fringe platforms. We note that the scope of the Act captures digital properties operated by mainstream media as well as fringe platforms and urge the Committee to consider how the Act is being applied in these contexts.

The role of regulation in combating illegal content

Google is supportive of regulation, where it is carefully crafted and appropriately tailored. A smart regulatory framework is essential to enabling an appropriate approach to illegal content. Four key principles inform our practises and we suggest that these provide a strong foundation for effective regulatory frameworks:

- **Shared Responsibility:** Tackling illegal content is a societal challenge—in which companies, governments, civil society, and users all have a role to play. It is essential to provide clear notice about the specific piece of content to an online platform, and then platforms have a responsibility to take appropriate action on the specific content. In some cases, content may not be clearly illegal, either because the facts are uncertain or because the legal outcome depends on a difficult balancing act; in turn, courts have an essential role to play in fact-finding and reaching legal conclusions on which platforms can rely.
- **Rule of law and creating legal clarity:** It's important to clearly define what platforms can do to fulfil their legal responsibilities, including removal obligations. An online platform that takes other voluntary steps to address illegal content will be reassured that these measures cannot have the negative consequences of being unprotected from legal liability. (This is sometimes called the "Good Samaritan" principle, and is reflected in leading legislative proposals such as the EU Commission's draft Digital Services Act regulation, which includes protections for "voluntary own-initiative investigations").
- **Flexibility to accommodate new technology:** While laws should accommodate relevant differences between platforms, given the fast-evolving nature of the sector, laws should be written in ways that address the underlying issue rather than focusing on existing technologies or mandating specific technological fixes.
- **Fairness and transparency:** Laws should support companies' ability to publish transparency reports about content removals, and provide people with notice and an ability to appeal removal of content. They should also recognise that fairness is a flexible and context-dependent notion—for example, improperly blocking newsworthy content or political expression could cause more harm than mistakenly blocking other types of content.

Specific feedback on the Act

Drawing on the above mentioned principles of smart regulation, we offer the following feedback on the Act;

1. ***We are concerned that the Act can be interpreted as requiring providers to proactively monitor all user generated content hosted on their platforms as it presumes providers to be reckless at the time that a notice is issued by the eSafety Commissioner, regardless of whether they were actually aware of the content.*** We support refinement of notice-and-takedown regimes, but we have significant concerns about laws that would mandate proactively monitoring or filtering content, impose overly rigid timelines for content removal, or otherwise impose harsh penalties even on those acting in good faith (noting that the Act does not include a Good Samaritan provision). These types of laws create a risk that platforms won't take a balanced approach to content removals, but instead take a "better safe than sorry" approach—blocking content at upload or implementing a "take down first, ask questions later (or never)" approach. We regularly receive overly broad removal requests, and analyses of cease-and-desist and takedown letters have found that many seek to remove potentially legitimate or protected speech. We note that the Attorney General's Department has included an assurance in their fact sheet that "The offence does not capture ignorance or negligence, and will not apply where a provider is genuinely unaware of particular material being accessible on their platform" and we would like to see the Act amended to make this more explicit. As an alternative, the presumption of recklessness could be removed from the Act to address this concern.

2. ***Time taken to remove / cease hosting abhorrent violent material should begin upon receipt of a notice from the eSafety Commissioner.*** We appreciate and support that the Act requires the removal of abhorrent violent material 'expeditiously' and would be concerned by any attempts to change this requirement to a specific number of hours. Setting a specific time period would not incentivise removal in the most expeditious manner, where removal for straightforward cases could be achieved before expiry of the specific time period. Our experience in implementing various frameworks elsewhere in the world that mandate a specific time for removal is that this does not allow for proper and appropriate review of more complex and borderline cases, which inevitably leads to overblocking of legitimate speech to avoid penalties for non-compliance.

3. ***The scope of services caught by the Act is overly broad for a criminal law and we request that search engines and services that allow end-users access to material primarily for business purposes be explicitly excluded from the definitions of designated internet services and hosting services within the Act.*** From a policy perspective, it's important to recognise the different purposes and functions of different services. Rules that make sense for social networks, video-sharing platforms,

and other services primarily designed to help people share content with a broad audience may not be appropriate for search engines and services used primarily for enterprise purposes, for which service providers may have limited abilities to remove content and where users have fundamentally different expectations and applications.

4. ***We would like to see more transparency from the Office of the eSafety Commissioner on decisions made and notices issued under the Act, as well as an appeals process through the Administrative Appeals Tribunal.*** We appreciate that for a notice issued under the Act to work effectively, it is important the eSafety Commissioner can act quickly and that this is why standard procedural fairness provisions were deliberately excluded from the operation of the Act. However, decisions taken by the Office of the eSafety Commissioner should be subject to public scrutiny and appeal after the fact, particularly in cases where the Commissioner's office has not contemplated the applicability of a defence under the Act. We also believe that service providers should be able to explain to the public why certain information has been removed from public circulation.

5. ***Kidnapping appears to be a lower order crime than the commission of murder, rape, torture or a terrorist act yet it is included within the definition of abhorrent violent content under the Act. The concept of kidnapping used in the Act is also broad and difficult to apply in practice. We suggest removing kidnapping from this definition.*** The concept of kidnapping used in the Act requires service providers to assess the purpose for which a person has been taken, which may be extremely difficult to determine. It also extends to kidnapping which is not violent, but where violence is threatened. Kidnapping without actual violence seems a significantly lower order crime than the other crimes included within the definition of abhorrent violent material. It is also difficult to determine, from the point of view of enforcing content moderation rules, whether a kidnapping is real or fake.

6. ***The obligation to notify the Australian Federal Police (AFP) within a reasonable time of becoming aware of the distribution of abhorrent violent material documenting conduct that has occurred or is occurring within Australia could present challenges where it is not apparent where the conduct is occurring.*** Service providers operating across multiple jurisdictions typically report such information to Interpol and rely on Interpol's expertise and networks to determine which law enforcement agency needs to receive the information. In instances where it is not apparent where the conduct is taking place, it would be more effective to allow notifications to the AFP or Interpol to assist in determining whether conduct is taking place within Australia. In addition, in the time since the Act was passed, industry has worked closely with the Department of Home Affairs to develop a Domestic Online Crisis Response Protocol that sets out the process to be followed when digital platforms become aware of an event that involves terrorist or extreme violent material being disseminated online in a manner likely to cause significant harm to the Australian community, and that warrants a rapid, coordinated and decisive response by industry

and relevant government agencies. We suggest that the obligation to notify the AFP under the Act be amended to include the option of notifying Interpol and to reflect the existence of this protocol.

7. ***The test of whether material is 'accessible' or 'hosted' would capture a situation where a perpetrator stores private videos on a site, without sharing them. This is arguably broader than the intended scope of the law (to prevent the widespread distribution of abhorrent violent material).*** We propose that Sections 474.34 (1)(b) & (5)(b) are amended so that it reads "the content service can be used for a person other than a person described in section 474.31 (1)(c) [ie a perpetrator or accomplice] to access material."

8. ***The research defence within the Act does not allow for publication of the research after it is conducted. We suggest that this defence be amended to avoid curbing public access to scientific, medical, academic or historical research.*** We suggest removing the words "but only where the accessibility is reasonable in the circumstances" and expanding the defence to include review and publication of the research for educational purposes.

9. ***The requirement within the journalism defence for material to be made by a "person working in a professional capacity as a journalist" is limiting and difficult to apply in practice.*** This requirement excludes valuable content produced by amateur journalists or concerned citizens. In our experience, it can be difficult to determine whether an individual is working in a 'professional capacity' as a journalist; would this include material written for authoritative news sources by guest contributors? We suggest that this limb of the defence be deleted.

10. ***The penalties under the Act are out of sync with other penalty regimes within the Australian criminal justice system.*** We appreciate that there is a spectrum of penalties that can be applied but would welcome explicit judicial guidance that the most significant penalties be reserved for bad faith actors and / or those who repeatedly and / or flagrantly breach the Act.

Please let us know if you have any follow up questions.

Yours faithfully

Samantha Yorke
Government Affairs and Public Policy