DIGI

Thursday April 18, 2019

Director, Online Content and eSafety Section
Department of Communications and the Arts
By email: onlinesafety@communications.gov.au

Dear Director,

Thank you for the extended opportunity to engage with the Australian Government about its proposed Online Safety Charter.

By way of background, the Digital Industry Group Inc. (DIGI) advocates for the interests of the digital industry in Australia. Its members include Google, Facebook, Twitter, Amazon and Verizon Media whose services range from search engines, content and communications platforms, and online stores. DIGI advocates for a balanced approach to technology policy that harnesses the tremendous social and economic opportunities digital services bring to Australia and globally, while also ensuring these services are used in a positive and beneficial way.

The recent tragic terrorist attacks in Christchurch have shed new light on the crucial importance of action against the dissemination of hate speech and violent content online, and provide important context for the discussion of frameworks such as the Online Safety Charter. No responsible Internet company wants to host such content, and they have a shared goal with governments in stopping its dissemination -- which means continually investing in people, technology and processes that ensure online spaces are safe.

DIGI looks forward to further engaging with the issues raised in the Online Safety Charter. Should you have any questions or wish to discuss any of the representations made in this submission further, please do not hesitate to contact me.

Best regards,

Sunita Bose
Managing Director
Digital Industry Group Inc. (DIGI)

The need for consultation

DIGI welcomes the opportunity to offer input to the Online Safety Charter, as consultation is crucially important to getting such decisions right. Unfortunately, recent regulation in relation to online safety has not received adequate consultation with affected and interested parties. Keeping the Internet safe is a crucially important goal we all share, but it is also not a straightforward challenge to solve. When the challenges of online safety are oversimplified, the laws proposed in response are ineffective, such as the amendment to the *Criminal Code on Abhorrent Violent Material* passed earlier this month. As this law was passed within the space of five days, it was not the subject of meaningful consultation with the relevant stakeholders -- including the technology industry, legal and technical experts, news media and civil society -- to ensure the solution proposed was effective. We urge the Government to consult closely with stakeholders in Australia and overseas to produce an online safety framework that addresses identified problems, and is in line with international efforts and norms in this area.

The need to address hate speech

In the wake of the tragic terrorist attacks in Christchurch, our political leaders highlighted the challenges of hate speech and extremism:

The Prime Minister Scott Morrison said shortly after the attacks:

> *"Extremism, or in a different form fundamentalism, is simply an inability to tolerate difference. It is to feel threatened by others who do not conform to your world view. And it takes many forms: religious extremism, secular extremism, and political extremism. Every terrorist attack has at its core a hatred of difference and a hatred about the choices and lives of others."*[1]

[1] Prime Minister Scott Morrison, 18/03/2019, "Australia-Israel Chamber of Commerce Speech", accessed at: https://www.pm.gov.au/media/australia-israel-chamber-commerce-speech

The Opposition Leader Bill Shorten said shortly after the attacks:

> *"Not all right wing extremist hate speech ends in right wing extremist violence. But all right wing extremist violence begins with right wing extremist hate speech."*[2]

DIGI is extremely concerned that the amendment to the *Criminal Code on Abhorrent Violent Material* -- the regulatory response conceived and passed in response to the terrorism recently experienced in Christchurch -- does not address hate speech, which is often a motivation for violent acts and terrorism. At the same time, Australia continues to adopt an archaic definition of hate speech under the Racial Discrimination Act (Cth) 1975 that only applies to race-based hate speech, and does not include religious-based or gender-based speech.

Frameworks to ensure online safety provide an opportunity to establish better foundations to combat hate speech. We therefore encourage the Australian Government to develop a clearer legislative framework that defines hate speech to assist enforcement agencies and prosecutors. This will also serve to help relevant stakeholders, including digital platforms, to better report, review and remove content that meets a defined Australian legal threshold.

As an example of a global initiative in this area, the European Commission developed the "Framework Decision on Combatting Racism and Xenophobia"[3] which criminalises the public incitement to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin. The EU, its Member States, digital services and other stakeholders, all share a collective responsibility to promote and safeguard freedom of expression in the online world and, at the same time, have a responsibility to combat violence and hatred. Hate speech, as defined in this Framework Decision, is a criminal offence also when it occurs online. To respond to the proliferation of racist and xenophobic hate speech online, the European Commission and four major digital companies (Facebook, Microsoft, Twitter and YouTube) presented a "Code of Conduct on countering illegal hate speech online" in May 2016[4]. Among other commitments under the European Code of Conduct, the industry has given a commitment under the code of conduct to remove the majority of illegal hate speech within 24 hours.

[2] Bill Shorten, 18/03/09, Tweet accessed at https://twitter.com/billshortenmp/status/1107450796642168832?lang=en

[3] European Commission, "Combating racism and xenophobia", accessed at https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/combating-racism-and-xenophobia_en

[4] European Commission, "Combating racism and xenophobia", accessed at https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/combating-racism-and-xenophobia_en

Additionally, the Australian Government may also provide further legal clarity by reviewing the protocol for listing terrorist organisations[5] in response to the growing threat from the far right and consider whether new organisations should be added. This might be similar to the FBI list of Foreign Terrorist Organisations and the UK's list of proscribed terrorist groups[6].

The need for global solutions

Ensuring online safety and disrupting the dissemination of hate speech and illegal content are global challenges that every country faces. We therefore need to ensure that Australia's approach to these challenges supports existing global approaches, and doing so has a number of benefits.

Firstly, looking at approaches that have been successful in other countries provides an opportunity to leverage existing expertise and learnings. For example, the European Commission recently announced that the fourth evaluation of its Code of Conduct on countering illegal hate speech online, described earlier, is delivering successful results. Specifically, the digital industry is "now assessing 89% of flagged content within 24 hours and 72% of the content deemed to be illegal hate speech is removed, compared to 40% and 28% respectively when the Code was first launched in 2016.[7]"

Secondly, global approaches help avoid conflicts of law, and fragmented country-by-country approaches. The recently passed amendment to the *Criminal Code on Abhorrent Violent Material* is a clear example of the dangers of a fragmented approach. There are laws in the United States, where all DIGI founding members are headquartered, that forbid companies from sharing certain types of information, specifically content data, with law enforcement agencies outside of the US. The obligation in Section 474.33 of this amendment to proactively share information with Australian law enforcement agencies creates a conflict of law. In this regard, DIGI members have long advocated for the introduction of the CLOUD Act which will enable the Australian and US Governments to enter a bilateral agreement that facilitates the lawful disclosure of content data to Australian law enforcement agencies. Such fragmented approaches to online safety, as demonstrated in this example, even risk Australia's important security co-operation relationship with the United States, and underline the importance of harmonising national approaches with global frameworks.

[5] Australian Government, "Listed terrorist organisations", accessed at https://www.nationalsecurity.gov.au/Listedterroristorganisations/Pages/default.aspx

[6] UK Government, 12/07/2013, "Proscribed terrorist groups or organisations", accessed at https://www.gov.uk/government/publications/proscribed-terror-groups-or-organisations--2

[7] European Commission, 4/2/2019, "Countering illegal hate speech online – EU Code of Conduct ensures swift response", accessed at http://europa.eu/rapid/press-release_IP-19-805_en.htm

The need for solutions across news media and digital services

It is important that Australia's approach to online safety engages all relevant players in the wider ecosystem. We believe further collaboration is appropriate between the digital industry, the telecommunications industry and the media industry. We would welcome the opportunity to liaise with the Communications Alliance to explore if and how an online safety framework might sit alongside codes adopted by the telecommunications industry. We also invite the Australian Communications & Media Authority (ACMA) to host a facilitated dialogue about the appropriate framework that can govern the removal of content established to be illegal across on mainstream media as well as digital services, similar to the process convened after the Christchurch terrorist attacks.

The need for transparency & accountability

We welcome opportunities to demonstrate how DIGI members can have processes for establishing accountability and transparency. Member companies continue to enhance transparency and accountability, both individually and in global partnerships, and several members publish transparency reports that identify the volume and types of content removed from their platforms, including government takedown requests. DIGI's founding members work cooperatively with the Office of the eSafety Commission on the removal of bullying content; there may also be an opportunity for the Office to publish more information on their complaint handling functions, including volumes and the platforms that are the subject of these complaints.

Relevant DIGI member work

DIGI founding members all release detailed information about their specific efforts in relation to online safety, and many members have released detailed statements pertaining to their response to the Christchurch attacks where relevant. To inform a better understanding of the status quo in relation to online safety efforts, outlined below is a high-level, brief summary of how DIGI members ensure online safety on their services.

Policies

Every DIGI member has policies outlining restricted content and user behaviour on their platforms, which are regularly updated to ensure they reflect emerging patterns of abuse. While policies vary on a product basis, at a high level, DIGI member services remove and restrict:

- Hate speech that attacks or maligns a group of people based on their protected class status;
- Content that promotes or glorifies violence;
- Bullying, harassment or abuse that directly threatens another person;
- Promotion of self-injury or suicide;
- Non-consensual sharing of intimate images;

- Child sexual exploitation;
- Various other illegal content in the markets where they operate.

Content moderation

The industry has also heavily invested in reporting tools and content moderation teams to ensure policy-violating content is surfaced and promptly actioned. DIGI's founding members maintain extensive review teams that operate to swiftly take appropriate action with user and community reports of policy-violating content. They also have expedited processes and protocols in place for urgent reports from law enforcement bodies, and for other content that requires rapid response. Reports of policy-violating and illegal content are reviewed and actioned by real people, who undergo extensive initial and ongoing training.

Technology

The industry has and continues to invest in technology to detect and prevent the dissemination of policy-violating content. This includes, but is not limited to:

- Image hashing, such as PhotoDNA to report and identify child sexual exploitation material.
- Machine learning algorithms that identify potentially problematic content before many people have consumed it and trigger a human review.
- A hash database that is shared amongst companies listing all known examples of terrorist content. At present this database includes almost 100,000 distinct pieces of content. Companies also coordinated within hours of the Christchurch terrorist attacks adding more than 1000 visually-distinct videos related to the attack to the collective hash sharing database. Crucially these were shared with smaller businesses that can benefit from this type of technology and information exchange.

Private & public sector collaboration

In addition to the sharing of online safety technology noted above, several DIGI founding members are pioneering a range of collaborative efforts across the industry, and with governments and with civil society, to address a wide range of issues related to online safety.

As one example, Google, Facebook, Microsoft, and Twitter launched the Global Internet Forum to Counter Terrorism (GIFCT) in 2017 with the objective to substantially disrupt terrorists' ability to promote terrorism, disseminate violent extremist propaganda, and exploit or glorify real-world acts of violence. The goals of the GIFCT are threefold: (i) building shared technology to prevent and disrupt the spread of terrorist content online, (ii) conducting and funding research by international experts, and (iii) sharing information and best practices with businesses of all sizes to assist them in managing this content on their platforms[8]. GIFCT has a growing community of partners, including companies

[8] Global Internet Forum to Counter Terrorism (GIFCT), https://www.gifct.org/about/

and civil society groups, and is moving to a new phase of development including a more formal membership model.

In Australia, DIGI has also welcomed the opportunity to work with the Australian Government to counter terrorism and extremism. We have participated in quarterly meetings with the Attorney General's Department and subsequently the Department of Home Affairs for several years, sharing information and identifying opportunities for collaboration. In addition, as part of this relationship, in November 2016 we hosted *DIGI Engage: our diverse digital future* together with the then Attorney-General's Department that brought together 150 people from around Australia and from diverse ethnic and religious backgrounds to promote tolerance, diversity and positive engagement online. In March 2018, DIGI and the Department of Home Affairs again partnered on this event, called *DIGI Engage 2018: Challenging Dangerous Online Narrative,* upskilling 100 young people from Australia and Southeast Asia came together in Sydney to combat extreme narratives online.

We will again host DIGI engage in 2019, at the request of the Department of Home Affairs. DIGI Engage 2019 will be a practical skills-building workshop for young leaders to explore the root causes of divisions in our societies and to workshop solutions such as counter-speech, in order to prevent and address hate and extremism. Alongside members' ongoing work to stop the dissemination of policy-violating and illegal content online, the partnership around the *DIGI Engage* events is an example of an important public-private collaboration that works to address and counter the societal causes of harmful content online -- recognising that the online world mirrors deeply complex and challenging societal problems that need to be addressed offline too.