# Submission to Parliamentary Joint Committee on Intelligence and Security

## Inquiry into extremist movements and radicalism in Australia

12 FEBRUARY 2020

FACEBOOK

## Executive summary

Facebook welcomes the opportunity to contribute to the Parliamentary Joint Committee on Intelligence and Security's (PJCIS') inquiry into extremist movements and radicalism in Australia. The terms of reference for this inquiry extend to "the role of social media platforms, encrypted communications platforms and the dark web in allowing extremists to communicate and organise".

Combatting terrorism and extremism is unfortunately a continuous responsibility for governments, working in partnership with experts, industry and the broader community. The existence of terrorist or extremist groups within society inevitably leads to terrorist or extremist activity online, and we take responsibility for detecting and removing these groups from Facebook's services.

Just as terrorist groups can quickly change tactics to evade detection in the real world, they also behave the same way online. We are constantly monitoring changes in their tactics and we have also been conscious of shifts in the terrorism threat environment throughout 2020. The Australian Security Intelligence Organisation has indicated "far-right extremism" and white supremacy now comprises up to 40 per cent of its workload.[1] And we have responded to the rapid rise and evolution in militarised social movements and a conspiracy theory that encourages violence, QAnon, including within Australia. This inquiry comes at a critical time and is an opportunity to make sure the coordinated response to terrorism threats in Australia remains relevant to the current threat environment.

In this submission, we outline the approach that Facebook takes to combatting hate and extremism on our services. We have significantly increased our commitments and investments in this area in recent years, and we now have 35,000 people working on safety and security within Facebook.

Our strategy comprises four elements:

1. **Policies**. Under our Community Standards, we have developed a number of policies that prevent hateful and extremist material on our services, including: (1) our dangerous individuals and organisations policy; (2) our policy on militarised social movements and violence-inducing conspiracy theories; and (3) our policies on hate speech and violence and incitement.

[1] M Truu, 'Threats from far-right extremists have skyrocketed in Australia, with ASIO comparing tactics to IS', *SBS News*, 22 September 2020, https://www.sbs.com.au/news/threats-from-far-right-extremists-have-skyrocketed-in-australia-with-asio-comparing-tactics-to-is

We regularly update our policies, in consultation with our community and relevant experts. In the last 12 months alone, we have made a number of important changes, such as:

- introducing a new 'hateful stereotypes' policy
- prohibiting any claims that deny or distort the Holocaust
- disallowing ads that claims that a group with "protected characteristics" is a threat to the safety, health or survival of others
- expanding our ads policies to better protect immigrants, migrants, refugees and asylum seekers from hateful claims.

2. **Enforcement.** We are continually taking steps to improve our ability to protectively detect hate and extremism on our services. We have banned more than 250 white supremacist organisations globally (a number which includes groups in Australia) and we have removed nearly 900 militarised social movements from our platform.[2]

   We recognise that we can always improve our enforcement, so we make data available to allow for scrutiny and accountability of the enforcement of our policies. We are increasingly identifying and removing violating content via artificial intelligence, so we don't need to rely on users seeing and reporting the content. In our last Community Standards Enforcement Report, we indicated that

   - 99.7 per cent of the terrorist content we removed was detected proactively
   - 97.5 per cent of the organised hate content we removed was detected proactively
   - 94.7 per cent of hate speech we removed was detected proactively.

3. **Partnerships**. While we have made significant progress as a company in combatting online hate and extremism, our work is enriched by partnerships with other companies, civil society organisations, experts, and governments. Some of our most important partnerships include:

   - the cross-industry group the Global Internet Forum to Counter Terrorism (GIFCT), of which we are a founding member. The GIFCT's database of shared digital "hashes" (fingerprints) and agreed protocols for responding to a live terrorist incident both improve our ability to enforce on our policies. The GIFCT Hash Sharing Database now contains approximately 300,000 hashes.[3]

[2] Facebook, 'An update to how we address movements and organizations tied to violence', *Facebook Newsroom*, blog post updated 19 January 2021, https://about.fb.com/news/2020/08/addressing-movements-and-organizations-tied-to-violence/.
[3] GIFCT, *GIFCT Transparency Report July 2020*, https://gifct.org/wp-content/uploads/2020/10/GIFCT-Transparency-Report-July-2020-Final.pdf.

- working with civil society groups to understand developments on the ground and to deploy programs to counter violent extremism. Initiatives like our Search Redirect Program or support for counterspeech initiatives help to combat radicalisation and push back against hate. We have also established an Australia-specific Combatting Online Hate Advisory Group, to ensure Australia civil society groups and experts have a direct channel to give us advice or feedback about how to better combat online hate.
- there is a significant amount of work we do in collaboration with governments and law enforcement and we contact law enforcement when we encounter credible threats of harm.

4. **Research**. We fund a significant amount of research to contribute to our own understanding of hate and extremism online, and to provide insights that contribute to the broader community of practice. In particular, we have commissioned two pieces of research specific to Australia to be publicly released in 2021: (1) hate speech experienced by Aboriginal and Torres Strait Islander people online; and (2) how LGBTQI+ Australians use our services, including how they combat online hate.

We encourage the PJCIS to consider not just how to prevent the violent manifestations of extremism, but also how to combat hate - as the root cause for extremism.

The terms of reference also cover the use of encrypted communications by terrorist and extremist groups. The PJCIS should acknowledge upfront that end-to-end encryption is the best security tool available to protect Australians from cybercriminals and hackers. However, it also poses a legitimate policy question: how to ensure the safety of Australians if no one can see the content of messages except the sender and the receiver?

The solution is for law enforcement and security agencies to collaborate with industry on developing even more safety mitigations and integrity tools for end-to-end encrypted services, especially when combined with the existing longstanding detection methods available to law enforcement. We already take action against a significant number of accounts on WhatsApp (a fully end-to-end encrypted messaging service) for terrorism reasons, and we believe this number could increase with greater collaboration from law enforcement and security agencies.

We also encourage the PJCIS to recommend full adoption of the INSLM report by the Government and amendment of the Assistance and Access Act in line with the INSLM's recommendations.

# Table of contents

# Facebook's approach to terrorist and extremist groups

Facebook has made significant commitments and investments to combat terrorist and extremist content on our platform. In particular, we now have more than 35,000 people working on safety and security within Facebook.

In this section, for the benefit of the PJCIS, we explain the approach that Facebook takes to combatting terrorism and extremism. Our strategy comprises:

1. Policies
2. Enforcement
3. Partnerships
4. Research.

## Policies

The policies that outline what is and is not allowed on Facebook are called our Community Standards.[4] Our policies are based on feedback from our community and the advice of experts in fields such as technology, public safety and human rights. Our Community Standards are also not static: we amend them regularly in response to feedback or developments.

A number of parts of our Community Standards are material to this inquiry, including our dangerous organisations policy, our policy on militarised social movements and violence-inducing conspiracy theories, our hate speech policy, and our policy on violence and incitement.

### Dangerous individuals and organisations

Facebook's Community Standards prohibit any organisation or individual that proclaims a violent mission or are engaged in violence from having a presence on Facebook. Specifically, we do not allow on our platform:

- terrorist organisations and terrorists
- hate organisations, and their leaders and prominent members
- mass / multiple murderers (including attempted murderers).

As well as removing these groups, we do not allow content that praises, supports or represents these groups.

Defining "terrorism" is a significant challenge. There is much debate among experts and policymakers about a definition of terrorism. It is a highly contested term, and

[4] Facebook, *Community Standards*, https://www.facebook.com/communitystandards/.

most governments or inter-governmental fora do not have an agreed term of terrorism.

However, as part of our industry-leading work to combat terrorism, Facebook has developed our definition of terrorism (which we use in assessing content on our platform). We define a terrorist organisation as:

> "Any non-governmental organization that engages in premeditated acts of violence against persons or property to intimidate a civilian population, government, or international organization in order to achieve a political, religious, or ideological aim."

Our definition is agnostic to the ideology or political goals of a group, which means it includes everything from religious extremists and violent separatists to white supremacists and militant environmental groups. It's about whether they use violence to pursue those goals. We have needed to develop a definition that can be applied consistently and equitably across the more than 3 billion people who use Facebook around the world.

### Militarised social movements and violence-inducing conspiracy theories

In August 2020, we expanded our dangerous organisations policy to capture "militarised social movements" and content relating to "violence-inducing conspiracy theories".

We have just begun to implement these policies, beginning with Pages, Groups, Events, and Instagram accounts dedicated to militarised social movements and violence-inducing conspiracy theories. Some examples of content that may be captured under this policy includes content relating to the violence at the US Capitol on 6 January 2021, such as militarised social movements like the Oathkeepers and a violence-inducing conspiracy theory like QAnon.[5]

### Hate speech

We don't allow hate speech on Facebook. It creates an environment of intimidation and exclusion, may promote offline violence, and can inhibit people from using their voice and feeling safe to connect freely.

[5] Facebook, 'An update to how we address movements and organizations tied to violence', *Facebook Newsroom*, blog post updated 19 January 2021, https://about.fb.com/news/2020/08/addressing-movements-and-organizations-tied-to-violence/.

We define hate speech as a direct attack against people on the basis of what we call protected characteristics. We have currently listed the following as protected characteristics:

- race
- ethnicity
- national origin
- disability
- religious affiliation
- caste
- sexual orientation
- sex
- gender identity
- serious disease.

We define attacks as violent or dehumanising speech, harmful stereotypes, statements of inferiority, expressions of contempt, disgust or dismissal, cursing, and calls for exclusion or segregation. This goes well beyond what is required in Australian legislation.

We have made a number of changes over the last 12 months to expand our hate speech policies in our Community Standards. These include:

- the development of a new hateful stereotypes policy, which will in the first instance prohibit content depicting blackface and stereotypes that Jewish people run the world.[6] We continue to consult on possible expansions to this policy to capture other hateful stereotypes.
- expansions in our ads policies to better protect immigrants, migrants, refugees and asylum seekers from hateful claims[7]
- expansions in our ads policies to prohibit claims that a group is a threat to the safety, health or survival of others on the basis of that group's race, ethnicity, national origin, religious affiliation, sexual orientation, gender, gender identity, serious disease or disability.[8]
- announcing that we will amend our policy to remove any claims that deny or distort the Holocaust, on the basis of expert consultation and research.[9]

[6] G Rosen, 'Community Standards Enforcement Report August 2020, *Facebook Newsroom*, 11 August 2020, https://about.fb.com/news/2020/08/community-standards-enforcement-report-aug-2020/.
[7] Facebook, 'Meeting the unique challenges of the 2020 elections', *Facebook Newsroom*, 26 June 2020, https://about.fb.com/news/2020/06/meeting-unique-elections-challenges/
[8] Ibid.
[9] M Bickert, 'Removing Holocaust denial content', *Facebook Newsroom*, 12 October 2020, https://about.fb.com/news/2020/10/removing-holocaust-denial-content//

### Violence and incitement

We aim to prevent potential offline harm that may be related to content on Facebook. While we understand that people commonly express disdain or disagreement by threatening or calling for violence in non-serious ways, we remove language that incites or facilitates serious violence. We remove content, disable accounts, and work with law enforcement when we believe there is a genuine risk of physical harm or direct threats to public safety. We also try to consider the language and context in order to distinguish casual statements from content that constitutes a credible threat to public or personal safety.

This policy means we are able to take action against content that is calling for violence or incitement, even if the author has not yet been designated by us as a dangerous organisation or individual.[10]

## Enforcement

Enforcing our policies against terrorist and extremist organisations is a constant challenge: just as terrorist groups have been resilient to counterterrorism efforts in the real world, we are in an adversarial situation in detecting and removing these groups. We need to continuously improve in order to help keep our community on Facebook safe.

Although our enforcement will not always be perfect, we have made significant progress in detecting and removing terrorist and extremist groups on our services. We have banned more than 250 white supremacist organisations globally and we have removed nearly 900 militarised social movements from our platform. Some of the individuals and organisations designed in Australia include Blair Cottrell, Neil Erickson, Tom Sewell, the Lads Society, the United Patriots Front, True Blue Crew and the Antipodean Resistance.

We detect dangerous organisations and terrorist content via a playbook and a series of automated techniques, which were first developed three years ago to detect content related to terrorist organisations such as ISIS, al Qaeda and their affiliates. We've since expanded these techniques substantially:

- We're now able to detect text embedded in images and videos in order to understand its full context

[10] As an example, see our work in relation to boogaloo content last year: Facebook, 'Banning a violent network in the US', *Facebook Newsroom*, 30 June 2020, https://about.fb.com/news/2020/06/banning-a-violent-network-in-the-us/.

- We've built media matching technology to find content that's identical or near-identical to photos, videos, text and even audio that we've already removed.
- We've now expanded to detect more groups tied to different hate-based and violent extremist ideologies and using different languages.
- We have learned from the techniques we currently use in the cyber security space to develop a new tactic that targets a banned group's presence across our apps. We do this by identifying signals that indicate a banned organisation has a presence, and then proactively investigating associated accounts, Pages and Groups before removing them all at once. Once we remove their presence, we work to identify attempts by the group to come back on our platform.
- We're also studying how dangerous organisations initially bypassed our detection, as well as how they attempt to return to Facebook after we remove their accounts, in order to strengthen our enforcement and create new barriers to keep them off our apps.
- We have been working to collect camera footage from law enforcement partners in the US and UK from their firearms training programs - providing a valuable source of data to train our systems. This should improve our detection of real-world, first-person shooter footage of violent events and avoid incorrectly detecting other types of footage. We have been collecting and ingesting that data from existing partners and hope to expand this collaboration to law enforcement agencies in other countries soon.
- We've increased our capability to rapidly respond to livestreams, including by reviewing all livestreams in an area that may involve footage of an attack and increasing our 24/7 capacity to respond to livestream reports.

In addition to building new tools, we've also employed new strategies, such as leveraging off-platform signals to identify dangerous content on Facebook, and implementing procedures to audit the accuracy of our artificial intelligence's decisions over time.

### Measuring the effectiveness of enforcement

We make data regularly available to assist in assessing and measuring the effectiveness of our enforcement approaches.

Our progress can be primarily measured through our transparent quarterly Community Standards Enforcement Report. We have long reported on the amount of terrorist content we have removed from our services, but for some time the reporting only covered content relating to Al Qaeda, ISIS and their affiliates. In 2019, we expanded our reporting to *al*l terrorist organisations; and, in 2020, we updated these

metrics to report on content that propagates organised hate (such as white supremacy) separate to terrorism content.[11]

According to the last Community Standards report (November 2020)[12], in the period July to September 2020, on Facebook, we took action against:

- 9.7 million pieces of content for terrorism
- 4.0 million pieces of content for organised hate
- 22.1 million pieces of content for hate speech.

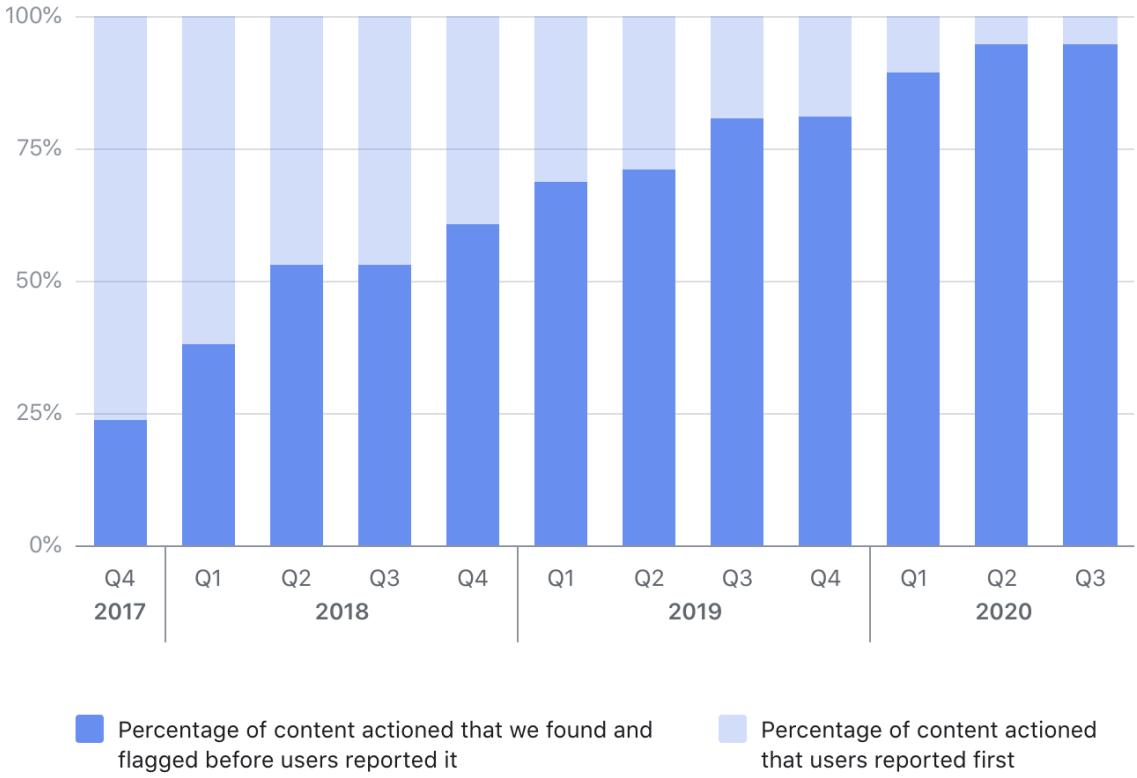
For each category, we also reported on the percentage of content that was detected proactively by us using artificial intelligence (compared to the percentage brought to our attention from a user report). Our ambition is to increasingly detect and remove content proactively, before users even see it, and so we have been investing significantly in artificial intelligence that helps us proactively detect this content. In the last reporting period:

- 99.7 per cent of the terrorist content we removed was detected proactively
- 97.5 per cent of the organised hate content we removed was detected proactively
- 94.7 per cent of hate speech we removed was detected proactively.

Our investment in artificial intelligence is evident from the increasing percentage of hate speech content we have been detecting proactively. Hate speech is traditionally one of the most challenging types of content to proactively detect because it is so context-dependent and challenging to develop and train artificial intelligence. At the end of 2017, less than 25 per cent of hate speech content we removed was detected proactively. This figure has progressively increased over that time: by end 2018, 40 per cent was proactively detected; by end 2019, 80 per cent was proactively detected (see **Figure 1** below). This improvement in our detection ability was accompanied by a stark increase in the total volume of hate speech content we have removed: at the end of 2018, we removed 3.4 million pieces of content; at the end of 2019, we removed 5.5 million; and in the last report in 2020, we removed 22.1 million.

[11] G Rosen, 'Community Standards Enforcement Report - August 2020', *Facebook Newsroom*, 11 August 2020, https://about.fb.com/news/2020/08/community-standards-enforcement-report-aug-2020/
[12] G Rosen, 'Community Standards Enforcement Report - November 2020', *Facebook Newsroom*, 19 November 2020, https://about.fb.com/news/2020/11/community-standards-enforcement-report-nov-2020/.

Figure 1: Hate speech removals on Facebook, by percentage of how they were detected




We have also developed a metric called *prevalence*, where we estimate how prevalent violating content is on Facebook. We think of this metric as how many views of violating content we did not prevent - either because people saw the content before we could take action, or because we missed the violation altogether.[13] We hold ourselves accountable to these numbers. In the last report:

- 0.10 to 0.11 per cent of views of content on Facebook contained hate speech. This means, for every 10,000 views of content on Facebook, 10 or 11 contained hate speech.
- For terrorist or organised hate content, there are insufficient views to precise estimate prevalence for these types of content. Because it is so infrequent, we estimate the upper limit for prevalence. For these types of content, it is 0.05 per cent of content views.

Our enforcement approach has been scrutinised externally. For example, a recent European Commission report found that Facebook assessed 95.7% and Instagram assessed 91.8% of hate speech notifications in less than 24 hours, compared to 81.5%

[13] A Kantor, 'Measuring our progress combating hate speech', *Facebook Newsroom*, 19 November 2020, https://about.fb.com/news/2020/11/measuring-progress-combating-hate-speech/.

for YouTube and 76.6% for Twitter.[14] The European Commission also stated that "only Facebook informs users systematically; all the other platforms have to make Improvements."

We will undergo an independent, third-party audit - starting this year - to validate the numbers we publish in our Community Standards Enforcement Report.[15]

## Partnerships

While we have made significant progress as a company in combatting online hate and extremism, our work is significantly enriched by partnerships with other companies, civil society organisations, experts, and governments. Some of these partnerships are outlined below.

### Cross-industry partnerships

Cross-industry partnerships are vital in countering online terrorism and extremism, because these groups generally work across multiple digital platforms and services to achieve their aims.

Facebook has been one of four founding members of a cross-industry partnership called the Global Internet Forum to Counter Terrorism (GIFCT).[16] It is a partnership that allows for collaboration and information-sharing to counter terrorism and extremism online, and works closely with governments, civil society and academia as well.

In 2020, the GIFCT transitioned to an independent organisation, appointed an inaugural and highly-respected Executive Director in Nicholas Rasmussen, and advanced significantly in the cooperative efforts implemented by its members. The GIFCT has also established an Independent Advisory Committee (which includes a NGO representative from Australia) and now includes a number of industry members.

The GIFCT has created a cross-industry database of "hashes" (unique digital fingerprints) of known violent terrorism imagery or propaganda. To date, the Hash Sharing Consortium has reached 300,000 unique hashes in the database - the result of approximately 250,000 visually distinct images and approximately 50,000 visually

[14] G Rosen, 'New EU report finds progress fighting hate speech', *Facebook Newsroom*, 23 June 2020, https://about.fb.com/news/2020/06/progress-fighting-hate-speech/.
[15] V Sarang, 'Independent audit of Community Standards Enforcement Report metrics', *Facebook Newsroom*, 11 August 2020, https://about.fb.com/news/2020/08/independent-audit-of-enforcement-report-metrics/.
[16]

distinct videos having been added. This helps to improve each company's ability to quickly detect and remove content involving a hash in the database.

The GIFCT has also developed a Content Incident Protocol - an agreed process for how companies will react if a real-world terrorist event triggers the sharing of online content. It was developed in response to the 2019 attacks in Christchurch.

### Civil society partnerships

Working with civil society organisations is critical to combatting hate and extremism. We regularly work with civil society organisations to hear feedback on our policies and enforcement, to understand trends and developments on the ground, and to reach memberships of the community at risk of radicalisation.

Some examples of our global partnerships include:

- Creation of a Search Redirect program. Search Redirect helps combat extremism by redirecting hate-related search terms on Facebook towards resources, education, and outreach groups. In 2019, we extended this program to Australia via a partnership with Exit Australia, a local organisation that helps people leave violent extremism and terrorism.
    - On International Holocaust Remembrance Day 2021, we launched a new Search Redirect module related to the Holocaust.[17] Anyone who searches on our platform for terms associated with either the Holocaust or Holocaust denial, will see a message from Facebook encouraging them to connect to the site www.aboutHolocaust.org which was created by the World Jewish Congress with the support of UNESCO (the United Nations Educational, Scientific and Cultural Organization) with the goal of providing people with essential information about the history of the Holocaust and its legacy.
    - We have also developed a Redirect initiative for QAnon. When someone searches for terms related to QAnon on Facebook and Instagram, we will redirect them to credible resources from the Global Network on Extremism and Technology (GNET), the academic research network of the GIFCT. These resources help inform people of the realities of QAnon and its ties to violence and real world harm.[18]
    - We have launched a similar Redirect Initiative for when people search for QAnon-adjacent terms related to child sex trafficking. When

[17] G Rosen, 'Connecting people to credible information about the Holocaust off Facebook', *Facebook Newsroom*, 27 January 2021, https://fb.workplace.com/groups/waitwhataskpr/permalink/5051911028190805/.
[18] Facebook, 'An update on our enforcement against QAnon', *Facebook Newsroom*, 21 October 2020, https://about.fb.com/news/2020/08/addressing-movements-and-organizations-tied-to-violence/.

searching for “save the children”, our prompt redirects users to the website of the actual NGO Save The Children.
    - Our Search Redirect initiative has been evaluated by Moonshot CVE as part of our commitment to ensuring the effectiveness of our program initiatives.[19]
- Counterspeech initiatives. One of the best methods for pushing back on hate speech is counterspeech: standing up to call out hate. Facebook works with NGOs around the world to support them in undertaking effective counterspeech, and we have created a hub[20] with resources and support specifically for NGOs.

Over the last twelve months, we have prioritised building partnerships with Australia-based organisations to assist in promoting counterspeech in Australian communities and to bring their specialised expertise to share trends about what they are seeing in Australian online communities. This engagement has taken a variety of forms, including

- undertaking concerted engagement with representatives from the Australian Jewish and Muslim communities to seek feedback on what they are seeing in relation to anti-Semitism and Islamophobia
- establishing an Australia-specific Combatting Online Hate Advisory Group in October 2020. The Advisory Group contains representatives of marginalised communities, and experts in different forms of online hate such as white supremacy. The Advisory Group has met twice and will continue quarterly meetings, to provide a forum to discuss how industry and civil society can work together closer in combatting online hate in Australia.

This builds on existing partnerships we have had within Australia, including a long-standing nine year partnership with PROJECT ROCKIT to help equip Australian school students with the skills required to engage online safely and push back on online hate.[21]

### Working with government and law enforcement

We also work closely with the Australian Government and other governments around the world on combatting terrorist and violent material. We have close ongoing engagement with law enforcement and security agencies. We have also instigated sessions with Australian law enforcement and security agencies to swap information on the terrorism threat environment within Australia.

---

[19] Moonshot CVE, *Facebook Redirect Programme: Moonshot Evaluation*, https://moonshotcve.com/facebook-redirect-programme-evaluation-report/
[20] Available at counterspeech.fb.com
[21] R Thomas, ‘Young people at the centre’, *Facebook Australia Blog*, 8 February 2021, https://australia.fb.com/post/young-people-at-the-centre.

Facebook was one of the signatories to the Christchurch Call, which was a ground-breaking commitment between multiple governments and technology companies led by the New Zealand Government.[22] We signed up to the voluntary nine-point industry plan, which contained a number of commitments to improve our effectiveness in combatting terrorist and extreme violent content.

We were also a member of the Australian Government Taskforce to Combat Terrorist and Extreme Violent Material Online, and we have been regularly reporting to the Australian Government on the Taskforce commitments since. This has included providing feedback to the Home Affairs Department in developing an Online Crisis Event Arrangement and participating in an Online Crisis Event simulation convened by the Department in October 2020.

We have worked with the Australian Government (and other governments) in international fora like the Organisation for Economic Cooperation and Development (OECD). There is significant work underway through the OECD on Voluntary Transparency Reporting Protocols, which was announced and sponsored by the Australian Government.[23] Facebook has been the only company to co-lead one of the working groups under this project; we have been co-leading a working group with the Australian Department of Home Affairs (previously the eSafety Commissioner's Office). We intend to continue to play an industry leadership role to support this important work through the OECD.

Finally, through our Australian industry association DIGI, Facebook has been working with the Department of Home Affairs on the annual event DIGI Engage, which is designed to empower young people to counter hate and extremism online.

## Research

In order to ensure our policies and enforcement approach reflects the latest research, we also partner with academics and experts.

Via the GIFCT, we have funded the Global Research Network on Terrorism and Technology (GRNTT) to develop research and provide policy recommendations around terrorists' and extremists' use of the internet. A total of 13 papers were

[22] Facebook, 'Facebook joins other tech companies to support the Christchurch Call to Action', *Facebook Newsroom*, 15 May 2019, https://about.fb.com/news/2019/05/christchurch-call-to-action/.
[23] S Morrison, *More action to prevent online terror*, media release 26 August 2019, https://www.pm.gov.au/media/more-action-prevent-online-terror.

produced and shared in openly accessible formats from the first phase of GRNTT's research.[24]

The second phase of GIFCT's research was led by the International Centre for the Study of Radicalisation (ICSR), based at King's College London. ICSR has established the Global Network on Extremism and Technology (GNET) and brings together an international consortium of leading academic institutions and experts with core institutional partnerships from the US, UK, Australia (The Lowy Institute), Germany and Singapore to study and share findings on combating terrorist and violent extremist use of digital platforms. The next phase of reports GIFCT has funded via GNET are in the process of being released.

These research reports are in addition to the insights reports that GNET publishes multiple times a week, which inform the work of GIFCT members.[25]

Facebook has also funded our own research round on misinformation and polarisation. 25 winners were announced in August 2020 and include two Australian proposals. A number of the successful proposals are examining polarisation (including how it can lead to extremism).[26]

We have also commissioned, funded or otherwise been involved with a number of other research reports relating to terrorism and extremism, including:

- The Centre for Analysis of the Radical Right has undertaken a report on A Guide to Online Radical-Right Symbols, Slogans and Slurs.[27] This includes symbols, slogans and slurs used by Australian members of the radical right.
- The Centre for Analysis of the Radical Right have also provided us a report on The Many Faces of the Radical Right and How to Counter Their Threat.[28]
- HOPE Not Hate have undertaken a report on the far right on Facebook[29]
- The Henry Jackson Society have delivered the report Free to Be Extreme[30]

[24] Global Network on Extremism and Technology, *Reports*, https://gnet-research.org/resources/reports/
[25] Global Network on Extremism and Technology, *Insights*, https://gnet-research.org/resources/insights/
[26] A Leavitt and K Grant, 'Announcing the winners of Facebook's request for proposals on misinformation and polarization', *Facebook Research Blog*, 7 August 2020, https://research.fb.com/blog/2020/08/announcing-the-winners-of-facebooks-request-for-proposals-on-misinformation-and-polarization/
[27] Centre for Analysis of the Radical Right, *A Guide to Online Radical-Right Symbols, Slogans and Slurs*, https://usercontent.one/wp/www.radicalrightanalysis.com/wp-content/uploads/2020/05/CARR-A-Guide-to-Online-Radical-Right-Symbols-Slogan-and-Slurs.pdf.
[28] Centre for Analysis of the Radical Right, *The Many Faces of the Radical Right and How to Counter Their Threat*, https://www.radicalrightanalysis.com/wp-content/uploads/2020/08/CARR-report-oD.pdf
[29] Hope Not Hate, *The Far Right on Facebook: a practical investigation into right-wing hate content on the platform*'.
[30] N Malik for the Henry Jackson Society, *Free to be extreme*, https://henryjacksonsociety.org/wp-content/uploads/2020/01/HJS-Free-to-be-Extreme-Report-FINAL-web.pdf

- Moonshot CVE has evaluated in a report the effectiveness of the Facebook Search Redirect program.[31]

We have also commissioned Australia-specific research to understand the experience of online hate from the perspective of two sets of potentially vulnerable groups:

- Aboriginal and Torres Strait Islander people. Research has been conducted by Dr Tristan Kennedy at Macquarie University is due to be released shortly.
- LGBTQI+ Australians. Research is being conducted by Dr Ben Hanckel from Western Sydney University and is also due to be released later this year.

We look forward to continuing to expand our efforts to fund research on hate and extremism in Australia and globally in 2021.

[31] Moonshot CVE, *Facebook Redirect Programme: Moonshot Evaluation*, https://moonshotcve.com/facebook-redirect-programme-evaluation-report/.

## Comments on encryption

The inquiry's terms of reference also relate to the use of encryption by terrorist actors.

It is critical to acknowledge upfront that end-to-end encryption is the best security tool available to protect Australians from cybercriminals and hackers. It is an essential component of cyber security and use of end-to-end encryption is so critical that it has become the global security standard for many online services, including private messaging services. All of the top ten messaging services in Australia (such as Apple's iMessage and Signal) offer end-to-end encrypted services. Taken in aggregate, end-to-end encryption is the norm today, not the exception, and people expect their messages to be safe.

However, it also poses a legitimate policy question: how to ensure the safety of Australians if noone can see the content of messages except the sender and the receiver?[32]

Some stakeholders are calling for the creation of a "backdoor" that would grant them power to read certain content. But it isn't that simple. Creating a backdoor requires building a structural weakness into a secure system used by billions of people every day. Once the weakness is there, we cannot choose who finds it. Cybercriminals are well resourced and technologically skilled: a backdoor for the good guys is just an open door for criminals. This is why Amnesty International has commented, "There is no middle ground: if law enforcement is allowed to circumvent encryption, then anybody can."[33]

UNICEF describes the debate around this issue well:

> *"End-to-end encryption is necessary to protect the privacy and security of all people using digital communication channels. This includes children, minority groups, dissidents and vulnerable communities. The UN Special Rapporteur on Freedom of Expression has referred to end-to-end encryption as "the most basic building block" for security on digital messaging apps. Encryption is also important for national security.*

[32] M Garlick, 'Online privacy and safety are not mutually exclusive', *Sydney Morning Herald*, 23 November 2020, https://www.smh.com.au/national/online-privacy-and-safety-are-not-mutually-exclusive-20201120-p56gml.html.
[33] Amnesty International, 'Government calls for Facebook to break encryption "latest attempt to intrude on private communications"', *Amnesty International News*, 4 October 2019, https://www.amnesty.org/en/latest/news/2019/10/government-calls-for-facebook-to-break-encryption-latest-attempt-to-intrude-on-private-communications/.

> *The debate around end-to-end encryption of digital communications has been polarized into absolutist positions. These include advocating 1) for the unlimited use of end-to-end encryption; 2) for the complete abolishment of end-to-end encryption; and 3) that law enforcement should always be able to access encrypted data and will be unable to protect the public unless it can do so. Such polarized positions ignore the complexity and nuance of the debate and act as an impediment to thoughtful policy responses. As noted by the Carnegie Endowment working group on encryption, polarized, absolutist positions in this debate should be rejected."*[34]

The solution is for law enforcement and security agencies to collaborate with industry on developing even more safety mitigations and integrity tools for end-to-end encrypted services, especially when combined with the existing longstanding detection and investigation methods available to law enforcement.

For our part, Facebook is committed to working with law enforcement, policymakers, experts and civil society organisations to develop ways of detecting bad actors without needing access to the content of encrypted messages.

We believe this approach will be far more effective than some of the alternatives mooted by some stakeholders, such as increasingly interventionist laws. In fact, the Independent National Security Legislation Monitor has already found that Australia's anti-encryption law, the Assistance and Access Act, is not "proportionate" nor appropriately protective of human rights.[35] The outcome of the INSLM's review (completed July 2019) is unknown, as the Government is yet to respond to the sensible and balanced recommendations put forward in the report *Trust But Verify*. As a result, the PJCIS has been unable to complete its review into the TOLA legislation, which was due in September 2020, and will be unable to fully consider the implications of the *Surveillance Legislation (Identify and Disrupt) Amendment*, which is also the subject of a separate PJCIS inquiry.

We encourage the PJCIS to recommend full adoption of the INSLM report by the Government and amendment of the Assistance and Access Act in line with the INSLM's recommendations.

---

[34] D Kardefelt-Winther, E Day, G Berman, S Witting and A Bose on behalf of the UNICEF cross-divisional task force on child online protection, *Encryption, Privacy and Children's Right to Protection from Harm*, https://www.unicef-irc.org/publications/pdf/Encryption_privacy_and_children%E2%80%99s_right_to_protection_from_harm.pdf

[35] Independent National Security Legislation Monitor, *Trust but verify*, https://www.inslm.gov.au/sites/default/files/2020-07/INSLM_Review_TOLA_related_matters.pdf.