

## Facebook responses to Questions of Notice - Inquiry on Foreign Interference through Social Media - Friday 30 July, 2021

### Question 1

**CHAIR, Senator Jenny McAllister:** Thank you for your opening statement and also for your written submission, which obviously was provided some time ago. Can I ask for an update about the prevalence of coordinated inauthentic behaviour on Facebook in the Australian context. Are you noticing particular trends in relation to coordinated inauthentic behaviour in Australia in relation to COVID?

**Mr Gleicher:** Coordinated inauthentic behaviour is only one type of threat—there are obviously others—but we generally see it as the most sophisticated tool used by determined adversaries like states and others. We've had four instances of CIB that have targeted Australia, in addition to other places, since 2017. We're happy to provide this afterwards, but I note that two months ago we published a threat report where we detailed all the lessons that we've learnt and the trends that we've seen in the context of influence operations between 2017 and 2020. We included with that a dataset of all 150-plus takedowns we conducted during that period, including the country of origin, the country of targeting, some public analysis and additional data and context. That could also be helpful here.

### Facebook response:

We consider authentic communications to be a central part of people's experience on Facebook. Our Community Standards, which are publicly available at [facebook.com/communitystandards](https://www.facebook.com/communitystandards), prohibit people engaging in inauthentic behaviour. This includes creating, managing, or otherwise perpetuating accounts that are fake, accounts that have fake names, and accounts that participate in, or claim to engage in, Coordinated Inauthentic Behaviour (CIB). We define Coordinated Inauthentic Behaviour as "any coordinated network of accounts, Pages and Groups that centrally relies on fake accounts to mislead Facebook and people using our services about who is behind it and what they're doing".<sup>1</sup>

---

<sup>1</sup> Facebook, 'Threat Report: The State of Influence Operations 2017-2020', *Facebook Newsroom*, <https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf>, May 2021.

We report regularly and transparently on our efforts to detect and remove CIB from our platforms. These are reported through:

1. **Our Community Standard Enforcement Report.** Each quarter, we report on metrics for preventing and taking action on content that goes against our Community Standards. Our Community Standards Enforcement Report is available at <https://transparency.facebook.com/community-standards-enforcement>
2. **Monthly CIB report.** Each month we publish a list of CIB networks that we have taken down. In some cases, we share information about the action taken at the time of enforcement. All reports and updates can be found at <https://about.fb.com/news/tag/coordinated-inauthentic-behavior/>
3. **Threat Report - State of Influence Operations 2017-2020.** This year we published a strategic report that looks at influence operations (IO) broadly, defined as “coordinated efforts to manipulate or corrupt public debate for a strategic goal”, of which CIB is a subset. The report draws on our existing public disclosures and our internal threat analysis to do four things: first, it defines how CIB manifests on our platform and beyond; second, it analyses the latest adversarial trends; third, it uses the US 2020 elections to examine how threat actors adapted in response to better detection and enforcement; and fourth, it offers mitigation strategies that we’ve seen to be effective against IO. The full report can be found at <https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf>

Since 2017, we have taken action on four instances of CIB operations in order to protect Australians, outlined below. We share these insights to ensure there is a strong, collective response to identifying and responding to threats that may occur across platforms.

- In 2020, we removed an operation that originated in China.<sup>2</sup> The operation used fake accounts to pose as locals in the countries they targeted including Australia, the Philippines, the United States and South East Asia more broadly. The operation posted in Groups, and commented and liked other people’s posts particularly related to naval activity in the South China Sea. The operation posted in Chinese, Filipino and English. We found this network as part of our internal investigation into suspected CIB in the region.

---

<sup>2</sup> N Gleicher, ‘Removing coordinated inauthentic behaviour in China and the Philippines’, *Facebook Newsroom*, <https://about.fb.com/news/2020/09/removing-coordinated-inauthentic-behavior-china-philippines/>, 22 September 2020.

- In 2019, we took action against CIB that originated in Macedonia and Kosovo.<sup>3</sup> The individuals behind this activity operated fake accounts to administer Pages sharing general, non-country specific content like astrology, celebrities and beauty tips. They also ran a small number of Pages purporting to represent political communities in Australia, the United Kingdom and the United States. Our investigation benefited from open source reporting, including from the press in Australia.
- In 2019, we took action against CIB that originated in the United Arab Emirates (UAE), Egypt, Nigeria that were sharing local news in targeted countries and promoting content about the UAE.<sup>4</sup> There were multiple sets of activity, each localised for a specific country or region, primarily in the Middle East and Africa, and some in Europe, North and South America, South Asia and East Asia, and Australia.
- In 2019, we took action against a domestic operation in March 2019 that was linked to local political actors related to the New South Wales state election.

---

<sup>3</sup> N Gleicher, 'Removing Coordinated Inauthentic Behaviour from Iran, Russia, Macedonia and Kosovo', *Facebook Newsroom*, <https://about.fb.com/news/2019/03/cib-iran-russia-macedonia-kosovo/>, 26 March 2020.

<sup>4</sup> N Gleicher, 'Removing Coordinated Behaviour in UAE, Nigeria, Indonesia and Egypt', *Facebook Newsroom*, <https://about.fb.com/news/2019/10/removing-coordinated-inauthentic-behavior-in-uae-nigeria-indonesia-and-egypt/>, 3 October 2019.

## **Question 2**

**CHAIR, Senator Jenny McAllister:** Thanks very much. I'm conscious that Facebook has been attempting to introduce some mechanisms to push people towards credible content during the pandemic. Your website indicates that two billion people have been connected to credible information about COVID-19 through your platform. Do you also track how much misinformation about COVID-19 has been seen on the platform? Do you have a similar data point about the bad stuff, I suppose, that's happening?

**Mr Gleicher:** My team is focused on behavioural deception—so not content that might violate our policies but behavioural tactics like CIB. So I don't have specific information on that, but I'd be happy to take that on notice and come back to you with more detail to make sure we can give you a comprehensive answer.

**Mr Machin:** There are probably a couple of points I could make here that might be helpful. We absolutely recognise that measuring misinformation is a real challenge, for a couple of reasons: first because people's views on whether a post on Facebook is misinformation or not can vary, and also because, particularly since the pandemic began last year, we've had to really rapidly scale up our policies and continue to consult with experts, and they have been shifting. A great example is that at the beginning of this year we had a policy that talking about blood clots as a side effect of COVID vaccines was harmful misinformation. Obviously then the science and advice we received from experts moved on. So it's very challenging. What we've done within Australia is recognise that additional information is needed to help people understand this particular policy issue and to understand whether the types of steps that we and others are taking are actually effective. So we've released some global statistics about the steps that we've taken to combat misinformation.

Since the beginning of the pandemic, we've removed harmful misinformation claims about COVID—18 million posts along those lines. We've worked with 80 fact checkers around the world, to whom we refer material that might not meet the threshold of leading to real-world harm, but we still want to have experts review it and let us know if it's false. There are 167 million posts that we've applied a 'false' label to after fact checkers have looked at them. Exactly as you say, Chair, we've also been trying to direct people to credible and authoritative sources of information. We've built a COVID information centre that we've been directly prompting Facebook and Instagram users to visit, and it has been visited by two billion people around the world since the pandemic began.

In order to help the debate within Australia, in the context of the industry code that Nathaniel mentioned earlier, we've prepared some Australia-specific versions of those statistics. This isn't something that we do very often, because it's very challenging, but there also are some limitations to how helpful it is because Australians who see a piece of information that comes from another country can still be affected by it. Measuring how much misinformation there might be within Australia doesn't necessarily give you the whole picture, but nonetheless, we recognise that we need to provide some more statistics in order to help that debate. What we have published is that in 2020 we removed 110,000 pieces of harmful COVID misinformation that originated in Australia. We've also provided information about the number of Australians who have visited our COVID information centre, which is 6.2 million people. It's quite significant in terms of the contribution we've been able to make with our services in directing people towards credible and authoritative information at a time when they most need it.

We know that measuring this information and providing some of these statistics is challenging but important. It's something that we're continuing to work on, so we hope that making those statistics available assists policymakers like yourselves in providing what we can at this point about the types of efforts that platforms like ours have been taking.

### **Facebook response:**

Throughout the pandemic, Facebook has been working hard to connect people to accurate, authoritative information, and remove harmful misinformation about COVID-19. We take a global approach to combatting misinformation, and we constantly update our efforts in response to feedback, research, and changes in the nature of misinformation. Our strategy to address misinformation on the platform has three parts:

First, Facebook removes misinformation that violates our Community Standards (available at [facebook.com/communitystandards](https://www.facebook.com/communitystandards)) and can cause imminent, physical harm. We work closely with experts around the world - in particular, the World Health Organization - to identify and update our policies on COVID-related claims that could cause imminent, physical harm. Our policies on COVID-19 are available at <https://www.facebook.com/help/230764881494641/>.

Second, we reduce access to information that has been rated as false by our third-party fact checking partners. Facebook partners with 80 third-party fact-checking organisations globally, and the Australian Associated Press and Agence France Presse in Australia, to assess the accuracy of content on our services. When they rate a piece of content as false, we add a prominent label warning people before they share it and show it lower in people's feed. Since 2016, Facebook has contributed more than \$84 million globally to support our fact-checking efforts.<sup>5</sup>

Third, we aim to inform people by providing access to up-to-date, authoritative health information. This includes but is not limited to providing contextual information around public posts, such as adding a context button or a breaking news tag, and providing a COVID-19 Information Centre with verified, authoritative information on COVID-19. The COVID-19 Information Centre is available at [https://www.facebook.com/coronavirus\\_info](https://www.facebook.com/coronavirus_info).

---

<sup>5</sup> C Alexander, 'Facebook launches accelerator challenge for global fact-checkers to expand reach of reliable information', *Facebook Journalism Project*, <https://www.facebook.com/journalismproject/accelerator-fact-checkers>, 26 August 2021.

This strategy has resulted in significant progress globally to promote reliable information and reduce harmful misinformation about COVID-19. We recently announced that since the start of the pandemic through to June 2021,<sup>6</sup> Facebook has:

- Removed more than 20 million pieces of content from Facebook and Instagram globally for violating our policies on COVID-19-related misinformation.
- We have removed over 3,000 accounts, pages, and groups for repeatedly violating our rules against spreading COVID-19 and vaccine misinformation.
- We displayed warnings on more than 190 million pieces of COVID-related content on Facebook that our third-party fact-checking partners rated as false, partly false, altered or missing context.
- We've provided authoritative information to help improve vaccine acceptance, connecting 2 billion people to resources from health experts through our COVID-19 Information Center.

Facebook has worked constructively with Government and industry in Australia to increase accountability and transparency around our misinformation efforts. In 2020, Facebook became a founding member and signatory to the Australian Disinformation and Misinformation Industry Code.<sup>7</sup> The code is a major step in establishing a regulatory framework around industry's work to combat misinformation and disinformation, with other countries around the world looking to emulate this approach.

Facebook has committed to 43 specific commitments to meet the obligations outlined in the voluntary code, and has begun reporting annually on our commitments (beginning with our first annual report in May 2021). This provides greater transparency to Australian policymakers and the community about the steps we take to combat disinformation and misinformation. Further detail on Facebook's commitments can be found at

<https://australia.fb.com/post/facebooks-response-to-australias-disinformation-and-misinformation-industry-code/>

---

<sup>6</sup> G Rosen, 'Community Standards Enforcement Report, Second Quarter 2021', *Facebook Newsroom*, <https://transparency.fb.com/data/community-standards-enforcement/>, 19 August 2021.

<sup>7</sup> J Machin, 'Facebook's response to Australia's disinformation and misinformation industry code', *Facebook Australia Blog*, <https://australia.fb.com/post/facebooks-response-to-australias-disinformation-and-misinformation-industry-code/>, 21 May 2021.

Importantly, as part of the code, we released Australia-specific statistics about content on our platform. Country-specific statistics about online content have limitations: they do not provide the full picture of what content Australians might see online. However, we recognise the importance of providing some data to contribute to a sophisticated public policy debate about misinformation in Australia, while industry and experts consider the best ways to measure and report on phenomena like online misinformation in the long term.

- Since the beginning of the pandemic in March 2020 to the end of December 2020, globally we removed over 14 million pieces of content that constituted misinformation related to COVID-19 that may lead to harm, such as content relating to fake preventative measures or exaggerated cures. 110,000 pieces of content were removed from Australia (noting that Australians benefitted from the content we removed from other countries as well).
- Since these figures were released, we have updated our global statistics on misinformation. Since the beginning of the pandemic, we have removed more than 20 million pieces of content from Facebook and Instagram globally for violating our policies on COVID-19 related misinformation. We have also displayed warning labels on more than 190 million pieces of COVID-related content that our third-party fact checking-partners rated as false, partly false, altered or missing context, to limit the spread of COVID-19 and vaccine misinformation.<sup>8</sup>
- We have made a COVID-19 Information Centre available around the world to promote authoritative information to Facebook users. More than 2 billion people globally, including over 6.2 million people in Australia visited the COVID-19 Information Centre during 2020.<sup>9</sup>

Our supplementary submission outlines some of the other ways we work to provide authoritative, credible information to Australians. We hope that the above information is helpful to the Committee, noting that we do not presently have data on how much misinformation that we have actioned has been seen.

---

<sup>8</sup> Facebook, *Community Standards Enforcement Report Q2 2021*, <https://transparency.fb.com/data/community-standards-enforcement/>

<sup>9</sup> J Machin, 'Facebook's response to Australia's disinformation and misinformation industry code', *Facebook Australia Blog*, <https://australia.fb.com/post/facebook-response-to-australias-disinformation-and-misinformation-industry-code/>, 21 May 2021.

### **Question 3**

**Senator VAN:** Thank you, gentlemen, for appearing before us. It's been a very useful session so far. I've only got about five minutes, so some of these questions I'll ask you to take on notice, including providing us with those reports that you mentioned. If there's a way for us to subscribe to your periodic releases of information could you inform us as to how to do that, because I'd like to keep on top of some of that.

Through the evidence we've heard previously, through both commercial and government operations, it has become clear that disinformation is only one part of foreign interference. I've put questions to other witnesses previously about the effects of foreign interference through social media. I hate to be alliterative, but they tend to fall under the Ds, as we call them: division, disinformation, distrust, destruction, deception and destabilisation. I'm particularly interested in the effects these have on social cohesion. Obviously electoral interference is terrible for a liberal democracy like ours and like that of the US, but I think some of what we're seeing is the breaking down of social cohesion in societies like Australia, the US and elsewhere. I'm interested in your views on how actors, whether state or otherwise, are using organic movements and then amplifying them to drive that breaking down of cohesion in civil society. I guess the starting point of that is—and you can take this on notice—what is your definition of 'foreign interference'?

**Mr Gleicher:** I'm happy for us to provide on notice a more detailed answer, but one thing that I would note is that foreign interference, from my perspective and from our perspective, covers a few different things. It could cover deceptive influence operations—what we call CIB. It could involve overt influence operations—for example, using labelled state media entities to push particular narratives. It could also include direct surveillance or hacking—cyberespionage. For example, we identified and removed a network not that long ago which was operating from China and targeting Uighur activists around the world. It was operating on Facebook, but it was actually much broader than that. Attempting to expose who they were and surveil them—you can imagine all the ways that would play into a puzzle like this.

There are two details that I would offer here that I think are important. The first is this. When we look at deceptive influence operations, CIB, about half of what we see is domestic in nature. While foreign interference is an important and very serious threat, we often see—as you described—foreign actors taking narratives from domestic actors, reflecting them or amplifying them, or we see entirely domestic operations using these techniques. In the conversation today, I think it is important to continue to focus on and tackle foreign interference, but we also have to think about what happens domestically. These are real citizens, from within our countries, that are driving these narratives. That tension is one that is going to continue to be a challenge for all of us, and we are seeing sophisticated foreign actors attempt to make themselves appear domestic and attempt to wander their narratives through domestic actors who may be sympathetic to their ideas.

**Mr Machin:** Senator Van, could I add one point which I think is helpful? Foreign interference aside, I think the impact of online discourse on social cohesion is something we do a lot of thinking about and we've done a lot of work on. I'm very happy to provide on notice the significant amount of work we've done to combat hate speech on our platforms, for example. This has traditionally been a real challenge for artificial intelligence and technology to proactively detect, but we're actually picking up more than 97 per cent of hate speech on Facebook before it's reported to us. It's an area where there's been a lot of



progress. Just drawing out the distinction between someone being hateful or intolerant—that can happen entirely independently of foreign interference. But, as Nathaniel says, we want to remove that content for a number of reasons, not least because we don't want it to be available to foreign actors to exploit. I just wanted to flag that, while we think about them separately, there are links. But there's a lot that we're doing in that space, and we'd be very happy to share it with you on notice.

### **Facebook response:**

There is a significant amount of work and investment that Facebook has made in ensuring the safety and security of our users, both online and offline. We now have more than 40,000 people working on safety and security at Facebook, and we've invested more than US\$13 billion (~AU\$17.5 billion) in safety and security since 2016. We expect that we will spend more than US\$5 billion (~AU\$6.6 billion) on safety and security in 2021 alone.

We committed to the Committee to provide more information on our work around social cohesion. Social cohesion spans across a number of our policies - it relates to our work on safety, hate speech, and violence and incitement. We have outlined below our continued efforts to support and strengthen social cohesion through our policies, enforcement efforts, partnerships and research.

### **Policies**

We set out the rules associated with what is and is not allowed on Facebook in our Community Standards, which are publicly available at [facebook.com/communitystandards](https://facebook.com/communitystandards). 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.

Four policies are of most relevance to the question of social cohesion. These are our policies on hate speech, violence and incitement, dangerous organisations and individuals, and militarised social movements and violence-inducing conspiracy theories.

- **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. We define hate speech as a direct attack against people on the basis of what we call protected characteristics including race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity, serious disease.

- **Violence and incitement.** 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.
- **Dangerous individuals and organisations:** Facebook prohibits 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; and, mass / multiple murderers (including attempted murderers). As well as removing these groups, we do not allow content that praises, supports or represents them.
- **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”. 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.

Even allowing for our policies, some have argued that polarisation has grown because of the influence of social media. This has been the subject of serious academic research in recent years. Many studies indicate polarisation has not been increasing in Australia and, in countries like the US where affective polarisation has increased, research suggests that social media is not the primary driver of polarisation. More details about this are discussed in a recent Medium post by our Vice President of Global Affairs Nick Clegg. Available here

<https://nickclegg.medium.com/you-and-the-algorithm-it-takes-two-to-tango-7722b19aa1c2>

One piece of research in particular covers Australia: Boxell, Gentzkow and Shapiro's article on Cross-Country Trends in Affective Polarisation.<sup>10</sup> These researchers examined trends in affective polarisation (the extent to which citizens feel more negatively towards other political parties than toward their own) across nine OECD countries over the past 40 years. Note: 'affective' polarisation is considered by experts to be more harmful than 'issue' polarisation, given that issue polarisation - disagreement on important issues - is not inherently bad and is in fact an essential part of the democratic political process. Affective polarisation is a better measure of whether citizens are able to "disagree well" with those who belong to different political parties.

The findings demonstrate that polarisation has remained stable, with a slight decrease in Australia since the mid-1990s.

### **Enforcement efforts**

We have invested heavily in artificial intelligence to proactively detect and remove content that violates our Community Standards, before a user necessarily needs to report it. We actively detect and action content that doesn't comply with our Community Standards across all Facebook accounts, Pages and Groups (both public and private), which means that even if someone doesn't report an issue to us, our artificial intelligence can detect potentially violating content and we can remove it.

Content such as hate speech has traditionally been the most challenging content for artificial intelligence to detect, because it is nuanced and context-dependent. However, due to the significant investments we have made in this area, our ability to proactively detect harmful content has increased significantly.

Our progress is evident through the Community Standards Enforcement Report, a voluntary transparency initiative that we release every quarter. In the latest report:<sup>11</sup>

- We removed 7.9 million pieces of content for bullying and harassment. 54.1 per cent was proactively detected via artificial intelligence. This is a significant increase from the beginning of 2019 when we removed 2.4 million pieces of content, 14.4 per cent of which were proactively detected.
- We removed 31.5 million pieces of content for hate speech. 97.6 per cent was proactively detected via artificial intelligence. Hate speech content removal has increased over 15X on Facebook and Instagram since we first began reporting it.
- We removed 6.2 million pieces of content for organised hate. 99.7 per cent was proactively detected via artificial intelligence. This is a significant increase from the beginning of 2020, when we first started reporting on this metric, when we removed 4.7 million pieces of content, 99.1 per cent of which were proactively detected.

---

<sup>10</sup> L Boxell, M Gentzkow & J Shapiro, Cross-Country Trends in Affective Polarization, <https://web.stanford.edu/~gentzkow/research/cross-polar.pdf>, June 2020.

<sup>11</sup> G Rosen, 'Community Standards Enforcement Report, Second Quarter 2021', Facebook Newsroom, <https://transparency.fb.com/data/community-standards-enforcement/>, 19 August 2021.

## Partnerships

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

- **Cross industry partnerships.** The Global Internet Forum to Counter Terrorism (GIFCT) is a cross-industry group of which we are a founding member. 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. 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.<sup>12</sup>
- **Civil social partnerships.** The Search Redirect Program 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.

Over the past twelve months, we have continued building partnerships with Australia-based organisations. These include ongoing 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.

We also established 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.

---

<sup>12</sup> GIFCT, *GIFCT Transparency Report July 2020*, <https://gifct.org/wp-content/uploads/2020/10/GIFCT-Transparency-Report-July-2020-Final.pdf>.

- **Government and law enforcement partnerships.** There is also 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. 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.<sup>13</sup> 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 are also working with the OECD on the Voluntary Transparency Reporting Protocols, which was announced and sponsored by the Australian Government.<sup>14</sup> We intend to continue to play an industry leadership role to support this important work through the OECD.

## Research

Finally, we fund a significant amount of research to contribute to our understanding of trends around online hate and extremism, and to provide insights that contribute to the broader community of practice.

Over the past two years we have invested over US\$4 million in a global round of funding for academic research on misinformation and polarisation across 2021 and 2022. Four research proposals from Australian universities were granted funding for their work. Proposals included ‘Testing fact and logic-based responses to polarising climate misinformation’ (John Cook and Sojung Kim, Monash University), ‘How fact checkers compare: News trust and COVID-19 information quality’ (Andrea Carson, James Meese, Justin B. Phillips, Leah Ruppner, La Trobe University)<sup>15</sup>, ‘Indigenous women and LGBTQI+ people and violence on Facebook’ (Bronwyn Carlson, Macquarie University), and ‘Unpacking trust and bias in social media news in developing countries’ (Denis Stukal, University of Sydney).<sup>16</sup>

---

<sup>13</sup> Facebook, ‘Facebook joins other tech companies to support the Christchurch Call to Action’, *Facebook Newsroom*, <https://about.fb.com/news/2019/05/christchurch-call-to-action/>, 15 May 2019.

<sup>14</sup> S Morrison, More action to prevent online terror, media release, <https://www.pm.gov.au/media/more-action-prevent-online-terror>, 26 August 2019.

<sup>15</sup> Facebook Research, ‘Announcing the 2021 recipients of research awards in misinformation and polarisation’, *Facebook Research*, <https://research.fb.com/blog/2020/08/announcing-the-winners-of-facebooks-request-for-proposals-on-misinformation-and-polarization/>, 14 September 2021.

<sup>16</sup> A Leavitt, K Grant, ‘Announcing the winners of Facebook’s request for proposals on misinformation and polarization’, *Facebook Research*, <https://research.fb.com/blog/2020/08/announcing-the-winners-of-facebooks-request-for-proposals-on-misinformation-and-polarization/>, 7 August 2020.

In Australia, we recently commissioned two pieces of research to understand the experience of online hate from the perspective of three sets of potentially vulnerable groups:

- Aboriginal and Torres Strait Islander people. Research was conducted by Dr Tristan Kennedy at Macquarie University.<sup>17</sup>
- LGBTQI+ Australians. Research is being conducted by Dr Ben Hanckel from Western Sydney University.<sup>18</sup>
- Asian Australians. The Online Hate Prevention Institute is reviewing and assessing whether Asian Australians have experienced more online hate since the COVID-19 pandemic.

---

<sup>17</sup> Dr B Hanckel, Dr S Chandra, 'Social media insights from sexuality and gender diverse young people during COVID-19', *Western Sydney University*, [https://www.westernsydney.edu.au/\\_data/assets/pdf\\_file/0006/1837977/Social\\_Media\\_and\\_LGBTQIA\\_Youth\\_Report.pdf](https://www.westernsydney.edu.au/_data/assets/pdf_file/0006/1837977/Social_Media_and_LGBTQIA_Youth_Report.pdf), May 2021.

<sup>18</sup> Dr T Kennedy, 'Indigenous peoples' experiences of harmful content on social media', *Macquarie University*, [https://research-management.mq.edu.au/ws/portalfiles/portal/135775224/MQU\\_HarmfulContentonSocialMedia\\_report\\_201202.pdf](https://research-management.mq.edu.au/ws/portalfiles/portal/135775224/MQU_HarmfulContentonSocialMedia_report_201202.pdf)

#### **Question 4**

**CHAIR:** Thanks, Mr Machin and Mr Gleicher. You mentioned in your testimony earlier that there were four occasions where Facebook identified CIB activity in Australia over the last couple of years. It would be of assistance to the committee if you could provide the details of those. I recognise that they are likely to have been published in a monthly report. Digging that out is not straightforward so it might be easier if you're able to provide that information to us directly.

#### **Facebook response**

Since 2017, we have taken action on four instances of CIB operations in order to protect Australians, outlined below. We share these insights to ensure there is a strong, collective response to identifying and responding to threats that may occur across platforms.

- In 2020, we removed an operation that originated in China.<sup>19</sup> The operation used fake accounts to pose as locals in the countries they targeted including Australia, the Philippines, the United States and South East Asia more broadly. The operation posted in Groups, and commented and liked other people's posts particularly related to naval activity in the South China Sea. The operation posted in Chinese, Filipino and English. We found this network as part of our internal investigation into suspected CIB in the region.
- In 2019, we took action against CIB that originated in Macedonia and Kosovo.<sup>20</sup> The individuals behind this activity operated fake accounts to administer Pages sharing general, non-country specific content like astrology, celebrities and beauty tips. They also ran a small number of Pages purporting to represent political communities in Australia, the United Kingdom and the United States. Our investigation benefited from open source reporting, including from the press in Australia.
- In 2019, we took action against CIB that originated in the United Arab Emirates (UAE), Egypt, Nigeria that were sharing local news in targeted countries and promoting content about the UAE.<sup>21</sup> There were multiple sets of activity, each localised for a specific country or region, primarily in the Middle East and Africa, and some in Europe, North and South America, South Asia and East Asia, and Australia.
- In 2019, we took action against a domestic operation in March 2019 that was linked to local political actors related to the New South Wales state election.

---

<sup>19</sup> N Gleicher, 'Removing coordinated inauthentic behaviour in China and the Philippines', *Facebook Newsroom*, <https://about.fb.com/news/2020/09/removing-coordinated-inauthentic-behavior-china-philippines/>, 22 September 2020.

<sup>20</sup> N Gleicher, 'Removing Coordinated Inauthentic Behaviour from Iran, Russia, Macedonia and Kosovo', *Facebook Newsroom*, <https://about.fb.com/news/2019/03/cib-iran-russia-macedonia-kosovo/>, 26 March 2020.

<sup>21</sup> N Gleicher, 'Removing Coordinated Behaviour in UAE, Nigeria, Indonesia and Egypt', *Facebook Newsroom*, <https://about.fb.com/news/2019/10/removing-coordinated-inauthentic-behavior-in-uae-nigeria-indonesia-and-egypt/>, 3 October 2019.