Submission to Australian Joint Standing Committee on Law Enforcement

Inquiry into law enforcement capabilities in relation to child exploitation

AUGUST 2021

FACEBOOK

Executive summary

Facebook welcomes the opportunity to provide a submission to the Joint Standing Committee on Law Enforcement's inquiry into *Law enforcement capabilities in relation to child exploitation*. The terms of reference for the inquiry refer to child sexual abuse material (CSAM) available on digital services and, given Facebook's significant efforts in combatting CSAM on our services, our submission is intended to assist the Joint Standing Committee by providing information about proactive steps being taken by industry.

Using the internet to harm children is abhorrent and unacceptable, and there is a continuous responsibility for all stakeholders - government, industry, and the broader community - to work together to protect children. Facebook has been an industry leader in initiatives to combat child exploitation, focussed especially on detecting, removing and reporting online CSAM on our services. Offenders can quickly change tactics to avoid detection, and we take responsibility for detecting and removing them from Facebook's services.

In this submission, we outline the approach that Facebook takes to protecting children 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 is based on four elements: developing **policies**, developing technology to **enforce** our policies by detecting and removing violating content, providing **tools** to support Australians to have a safe and positive experience on our services, and establishing **partnerships** with NGOs, other digital platforms and governments to encourage collaboration in protecting children online.

The impact of our efforts is clear: in the last quarter alone, we removed 25.7 million pieces of content for child sexual exploitation, and 99.5% of this content was detected and removed by us proactively before a user needed to see it and report it to us.[1] When we detect CSAM, we report it to the non-government organisation (NGO) the National Center for Missing and Exploited Children (NCMEC), a nonprofit that refers cases to law enforcement in Australia and around the world, in compliance with US law.

The technology we have invested in to detect and remove CSAM is cutting edge. For example, we have developed two technologies (called PDQ and TMK+PDQF) to detect identical and near-identical photos and videos -- and we have made these technologies available open source for free to allow industry partners, small developers and NGOs to benefit from this technology too. The President and CEO of NCMEC John Clarke said, "We're confident that Facebook's generous contribution of this open-source technology will ultimately lead to the identification and rescue of

[1] Facebook, *Community Standards Enforcement Report Q2 2021 - Child nudity and sexual exploitation*, https://transparency.fb.com/data/community-standards-enforcement/child-nudity-and-sexual-exploitation/facebook/.

more child sexual abuse victims."[2] The Australian Federal Police have reviewed these algorithms and are now using them as part of their work to protect children within Australia.[3]

We're going even further to develop tools to prevent inappropriate interactions between adults and minors on our services. We recently announced that we are identifying those accounts that exhibit potentially suspicious behaviour and stopping those people from interacting with young people.[4] Australia is one of the first countries in the world where we are rolling out this capability.

We are continuing to apply this type of innovative thinking as technology evolves. For example, we know that end-to-end encryption provides the strongest possible protection from cybersecurity threats and has become the industry standard for many applications, including private messaging. However, encryption poses legitimate policy questions about how to protect the safety of users if only the recipient sees the content of private messages. The type of technology that we have developed around inappropriate interactions with young people works without needing to see the content of private messages, demonstrating that there continues to be significant innovation in how to combat CSAM online.

The relationship between technology companies and law enforcement continues to be essential to stopping offenders from abusing our services, and we look forward to opportunities to deepen that engagement further.

[2] A Davis and G Rosen, 'Open-Sourcing Photo- and Video-Matching Technology to Make the Internet Safer', *Facebook Newsroom*, 1 August 2019, https://about.fb.com/news/2019/08/open-source-photo-video-matching/.
[3] J Dalins, C Wilson and D Boudry, 'PDQ & TMK+PDQF - A test drive of Facebook's perceptual hashing algorithms', *Journal of Digital Investigations*, pre-print, submitted December 2019.
[4] Facebook, 'Giving young people a safer, more private experience on Instagram', *Facebook Newsroom*, 27 July 2021, https://about.fb.com/news/2021/07/instagram-safe-and-private-for-young-people/.

Table of contents

Introduction

Using our apps to harm children is abhorrent and unacceptable. Facebook thinks about our efforts to combat online CSAM as falling in three areas:

- We want to **prevent** abuse of our services in this way in the first place.
- If an offender circumvents our prevention efforts, we want to **detect** that content or behaviour.
- Once abuse is detected, we **respond** by reporting that material to NCMEC, which provides it onto law enforcement.

We've also undertaken in-depth analysis to understand how and why people share child exploitative content on Facebook's services. As part of this, we analysed a sample of our reports to NCMEC and found that most of the CSAM on our services were copies of known material:

- More than 90% of our NCMEC reports were the same or visually similar to material that had been previously reported to NCMEC.
- More than half of the child exploitative content we reported were copies of just **six** videos.[5]

From this, we can understand that the number of pieces of CSAM content does not equal the number of victims. Instead, this suggests the behaviour that occurs on our services is largely revictimisation of the same victim by repeatedly sharing the same content.

Whilst every sharing of child exploitative content is inexcusable and harmful, analysing the nature of our reports assists us to identify that, to effectively stop this sharing of CSAM, we need to understand the intent behind the sharers. We worked with global child safety experts - including NCMEC - to develop a taxonomy of people's intent in sharing this material, based on existing research.[6] People who share these images are not a homogenous group; there are a variety of intentions. As well as those who have malicious intent towards children, people may share CSAM with nonmalicious intent (for example, out of shock or outrage, out of ignorance, in poor humour [eg. someone sharing an image of a child's genitals being bitten by an animal], or children sending sexual imagery of themselves to another child). While our work to understand intent is still ongoing, our initial estimates suggest that 75 per cent of CSAM sharing on our services is due to people sharing it with non-malicious intent.

It is analysis and understanding like this that has informed the comprehensive approach we've taken to combatting child exploitation and sharing of CSAM on our services.

[5] A Davis, 'Preventing child exploitation on our apps', *Facebook Newsroom*, 23 February 2021, https://about.fb.com/news/2021/02/preventing-child-exploitation-on-our-apps/ .
[6] J Buckley, M Andrus and C Williams, 'Understanding the intentions of Child Sexual Abuse Material (CSAM) sharers;, *Facebook Research Blog*, 23 February 2021, https://research.fb.com/blog/2021/02/understanding-the-intentions-of-child-sexual-abuse-material-csam-sharers/.

Facebook's work in combatting child exploitation

Our work to combat child exploitation falls in four broad categories:

1. **Policies** to set out what material is and is not allowed on our services
2. **Enforcement** of those policies via advanced technology
3. **Tools** to support Australians to have a safe and positive experience on our services
4. **Partnerships** with NGOs, other digital platforms, and governments to encourage collaboration in protecting children online.

Each of these is outlined in more detail below.

Policies

The policies about what material is and is not allowed on Facebook is contained in our Community Standards.[7] We have long had a very clear policy that CSAM is not permitted on our services. This policy is broader than just material that depicts sexual intercourse; we also do not allow:

- Child nudity
- Content that involves a child and includes sexual elements (for example, restraints, a focus on genitals, presence of an aroused adult, presence of sex toys, sexualised costumes, stripping, a staged environment or professionally shot, or open-mouth kissing)
- Content of children in a sexual fetish context
- Content that supports, promotes, advocates or encourages participation in paedophilia
- Content that identifies or mocks alleged victims of child sexual exploitation by name or image
- Solicitation content (for example, soliciting imagery of child sexual exploitation or real-world sexual encounters with children)
- Content that constitutes or facilitates inappropriate interactions with children (for example, engaging in implicitly sexual conversation with children or obtaining or requesting sexual material from children).

There are adjacent types of content that we also do not allow on our services. For example, in 2020, we expanded our policies to prohibit the implicit sexualisation of minors (in addition to our pre-existing policies against the explicit sexualisation of children). This can be a challenging category of material to detect that requires fine judgements to be made: for example, a user who comments on a benign photo of a child by saying it is "beautiful" could be, depending on the context, either providing an innocent compliment or inappropriately sexualising a child.

We also restrict the display of nudity or sexual activity of adults more generally, and content that involves the non-sexual abuse of children.

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

These policies are developed in close consultation with global experts, including in Australia. We convene a global Safety Advisory Board (which contains Australian experts, like PROJECT ROCKIT), quarterly virtual roundtables with Australian stakeholders, and specific consultation with subject matter experts when we're considering potential policy changes.

Enforcement

In order to enforce our policies, we investigate very significantly in both technology and people to help detect violating content, or suspicious behaviour.

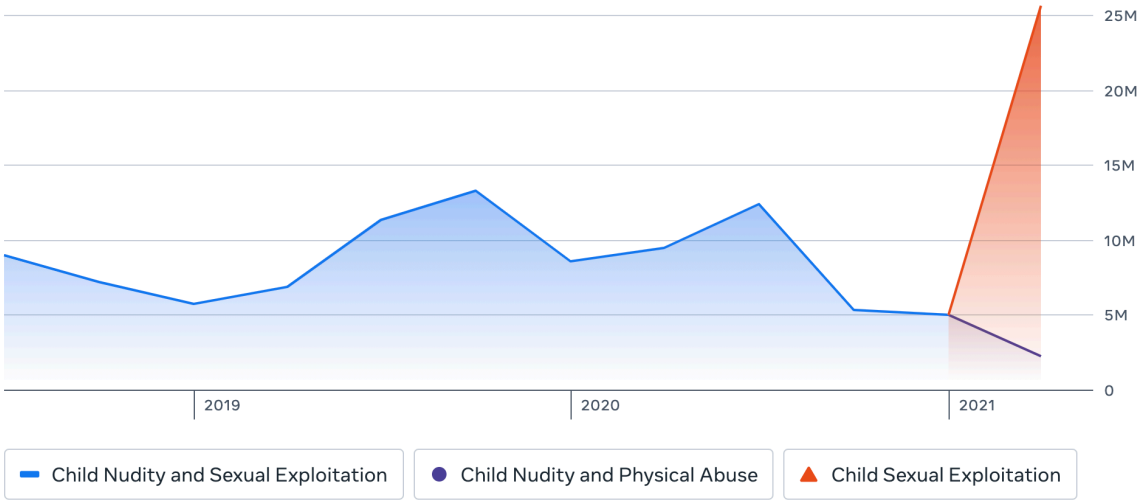
Firstly, we build up teams of experts who work in this space. The number of people working on safety and security has increased to more than 35,000 in recent years.

Secondly, the technology we have invested in to detect and remove CSAM is cutting edge. For example, we have developed two technologies (called PDQ and TMK+PDQF) to detect identical and near-identical photos and videos -- and we have made these technologies available open source for free to allow industry partners, small developers and NGOs to benefit from this technology too. The President and CEO of NCMEC John Clarke said, "We're confident that Facebook's generous contribution of this open-source technology will ultimately lead to the identification and rescue of more child sexual abuse victims."[8] The Australian Federal Police have reviewed these algorithms and are now using them as part of their work to protect children within Australia. We use these technologies along with many other examples of artificial intelligence.

Our work has a significant impact. In the last quarter alone, we removed 25.7 million pieces of content for child sexual exploitation, and 99.5% of this content was detected and removed by us proactively before a user needed to see it and report it to us.[9] For many years, we have detected millions of pieces of CSAM, consistently more than 99% detected proactively by us before users report it to us, which requires them to have seen it first.

[8] A Davis and G Rosen, 'Open-Sourcing Photo- and Video-Matching Technology to Make the Internet Safer', *Facebook Newsroom*, 1 August 2019, https://about.fb.com/news/2019/08/open-source-photo-video-matching/.
[9] Facebook, *Community Standards Enforcement Report Q2 2021*, https://transparency.fb.com/data/community-standards-enforcement/child-nudity-and-sexual-exploitation/facebook/.

Graph 1: Volume of child endangerment content detected and removed from Facebook 2018-2021



Note: From Q2 2021, we have broken out and reported separately on child sexual exploitation versus child nudity. Prior to this time, both categories of content were reported together.

When we become aware of CSAM, we report it to the NGO the National Center for Missing and Exploited Children (NCMEC), a nonprofit that refers cases to law enforcement in Australia and around the world, in compliance with US law. Facebook works closely with NCMEC to improve the ecosystem to fight this abuse, for example, by recently rebuilding their case management tool pro-bono in order to provide greater context around a particular report when it is provided to law enforcement around the world.

Tools

We offer a number of tools in this space, including:

- tools for parents and young people to support them in having a safe experience on our services.
- tools to deter potential offenders.

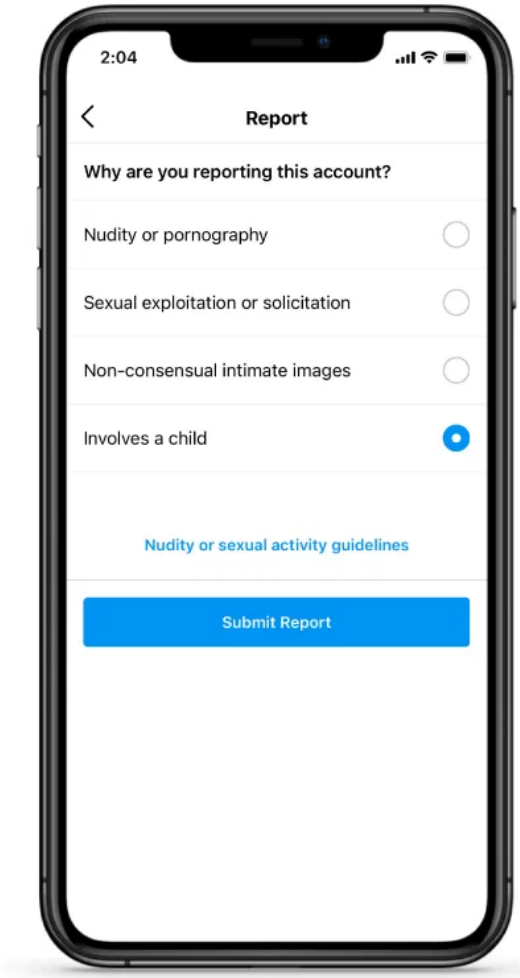
Tools for parents and young people

We have longstanding tools that young people can take in order to protect the privacy of their own accounts, including limiting who can find them, who can send them a friend request and what information is publicly available. We've also provided longstanding options to Block, Report, Hide or Unfollow users.

We want to stop young people from hearing from adults they don't know or don't want to hear from, and we believe private accounts are the best way to prevent this from happening. Since July 2021, everyone who is under 16 years old in Australia is defaulted into a private account when they join Instagram. For young people who already have a public account on Instagram, we'll show them a notification

highlighting the benefits of a private account and explaining how to change their privacy settings.[10] We have also been investing significantly in artificial intelligence in order to detect the age of young users, especially those who may be under 13 and too young to use our apps.[11]

After consultations with child safety experts and organisations, we've made it easier to report content for violating our child exploitation policies. To do this, we added the option to choose "involves a child" under the "Nudity & Sexual Activity" category of reporting in more places on Facebook and Instagram. These reports are prioritised for review.



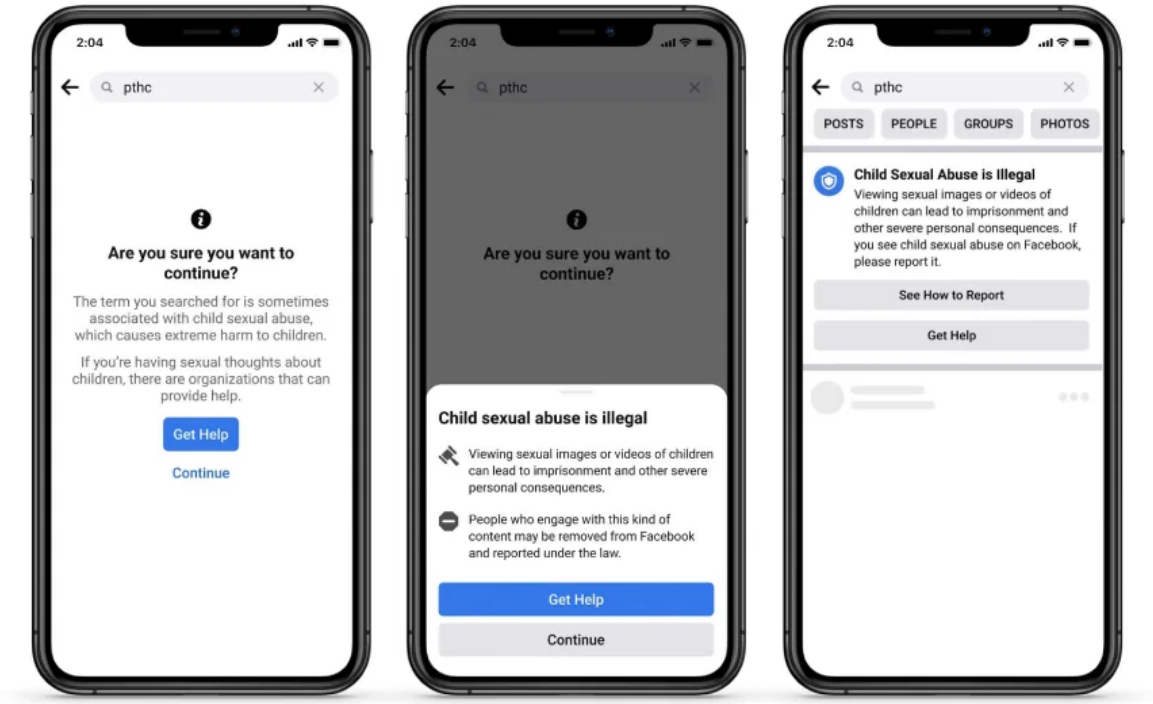
We've built a hub in our Safety Centre, dedicated to helping parents understand the various tools available to protect the safety of young people on our services. It can be accessed at www.facebook.com/safety/childsafety.

[10] Facebook, 'Giving young people a safer, more private experience on Instagram', *Facebook Newsroom*, 27 July 2021, https://about.fb.com/news/2021/07/instagram-safe-and-private-for-young-people/.
[11] P Diwanji, 'How do we know someone is old enough to use our apps?', *Facebook Newsroom*, 27 July 2021, https://about.fb.com/news/2021/07/age-verification/.

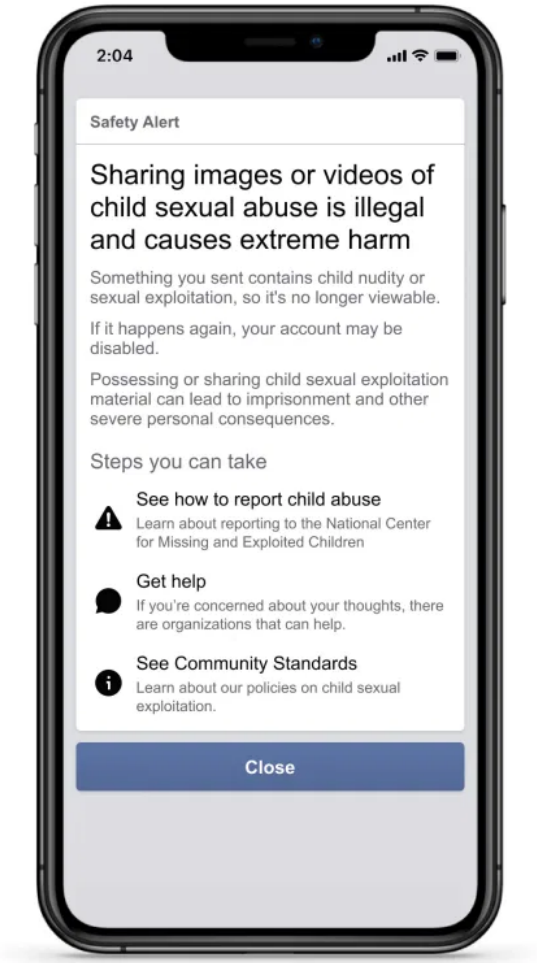
Tools to deter potential offenders

Based on our research and analysis about users with potentially malicious vs non-malicious intent, we have a range of customised interventions for users who may be looking for CSAM on our services.

We've started by testing two new tools — one aimed at the potentially malicious searching for this content and another aimed at the non-malicious sharing of this content. The first is a pop-up that is shown to people who search for terms on our apps associated with child exploitation. The pop-up offers ways to get help from offender diversion organisations and shares information about the consequences of viewing illegal content.



The second is a safety alert that informs people who have shared viral, meme child exploitative content about the harm it can cause and warns that it is against our policies and there are legal consequences for sharing this material. We share this safety alert in addition to removing the content, banking it and reporting it to NCMEC. Accounts that promote this content will be removed. We are using insights from this safety alert to help us identify behavioural signals of those who might be at risk of sharing this material, so we can also educate them on why it is harmful and encourage them not to share it on any surface — public or private.

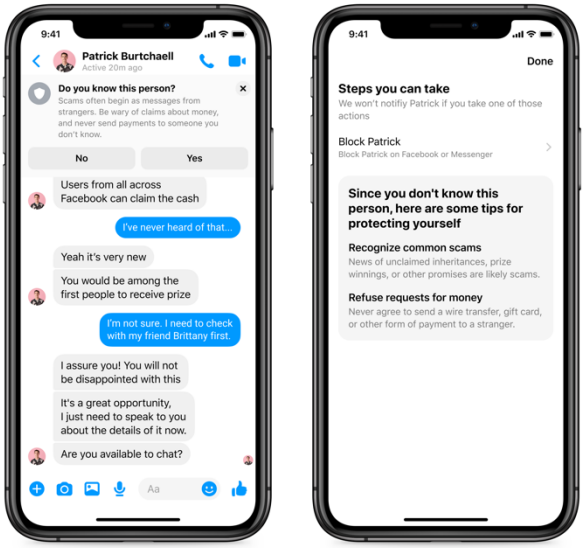


We are also taking steps to make it harder for potential suspicious accounts to contact young users. We've developed new technology that will allow us to find accounts that have shown potentially suspicious behaviour and stop those accounts from interacting with young people's accounts. By "potentially suspicious behaviour", we mean accounts belonging to adults that may have recently been blocked or reported by a young person, for example.

Using this technology, we prevent young people's accounts from appearing in recommendations to these adults. If they find young people's accounts by searching for their usernames, they aren't able to follow them. They aren't able to see

comments from young people on other people's posts, nor will they be able to leave comments on young people's posts.

Since 2020, we have also sent notices to users in Messenger where we believe an adult could be pursuing a potentially inappropriate private interaction with a child. These are used in instances where someone may be grooming or scamming another user.[12]



Partnerships

While we undertake a lot of work to ensure our own services are safe, we know that online CSAM is an industry-wide problem and requires collaboration between digital platforms and governments, law enforcement, safety experts, NGOs and parents. It's our collective responsibility to combat abuse and protect young people online.

In 2020, we joined with Google, Microsoft and 15 other tech companies to announce the formation of "Project Protect: A plan to combat online child sexual abuse", a renewed commitment and investment from the Technology Coalition expanding its scope and impact to protect kids online and guide its work for the next 15 years.

Project Protect is focussing on five key areas:

- **Tech innovation:** Accelerating the development and usage of groundbreaking technology. All companies involved have contributed to a multi-million dollar fund to support this work
- **Collective action:** Convening tech companies, governments and civil society to create a holistic approach to tackle this issue
- **Independent research:** Funding research with the End Violence Against Children Partnership to advance our collective understanding of the

[12] J Sullivan, 'Preventing unwanted contacts and scams in Messenger', *Messenger News*, 21 May 2020, https://messengernews.fb.com/2020/05/21/preventing-unwanted-contacts-and-scams-in-messenger/.

experiences and patterns of child sexual exploitation and abuse online, and learn from effective efforts to prevent, deter and combat it
- **Information and knowledge sharing:** Enabling greater information, expertise and knowledge sharing among companies to help prevent and disrupt child sexual exploitation and abuse online
- **Transparency and accountability:** Increasing accountability and consistency across the industry through meaningful reporting of child sexual exploitation and abuse content across member platforms and services. This will be done in conjunction with WePROTECT Global Alliance.

We also work closely with a range of Australian child safety NGOs, to ensure we are able to review and consider any material they provide us.

Based on the analysis we've undertaken of the intent behind sharing of CSAM (discussed earlier), we will shortly be launching a public service announcement in a number of markets, including in Australia around National Child Protection Week. The PSA will spread the message of "report it, don't share it" in order to educate members of the community who may fall into the category of non-malicious sharers of CSAM. The PSA will direct Australians to report material they see either to Facebook or to the Office of the eSafety Commissioner.

We also work with Australian law enforcement in a variety of ways, including by helping to amplify their public service announcements. The Australian Centre for Child Exploitation has had great success using Facebook and Instagram for their recent series, which reached over 1.3 million people and increased traffic to their website 1110%. They said "Social media is having a tremendous impact in both the prevention and operational work of the ACCCE, and we thank our loyal followers and partners who are working with us to fight online child sexual exploitation and win."[13] We have also been a key sponsor for many years of Taskforce Argos' Youth Technologies and Virtual Communities Conference, a globally recognised forum that supports practitioners in the fields of law enforcement, prosecution, education, child protective services, social work, children's advocacy and therapy who work directly with child victims of crime.

We also have a range of additional partnerships to assist with education and empowerment of children and parents engaging online:

- PROJECT ROCKIT's Digital Ambassadors program, in which Facebook has invested more than $1 million. Digital Ambassadors is a youth-led, peer-based anti-bullying initiative. A Digital Ambassador aims to utilise credible strategies to safely connect and tackle online hate. We are supporting PROJECT ROCKIT to continue to reach young people through education, particularly in remote and regional areas throughout 2021. This is a nine-year partnership that has directly empowered more than 11,500 young Australians to tackle

[13] Australian Centre to Counter Child Exploitation, *Newsletter June 2021*, https://www.accce.gov.au/news-and-media/newsletter/newsletter-june-2021.

cyberbullying.[14] The most recent version of the program has been launched by the eSafety Commissioner herself.

Australian eSafety Commissioner launching the virtual version of Digital Ambassadors

- We worked with the Alannah and Madeline Foundation and the Stars Foundation on the Safe Sistas program, which supports the online safety of young Indigenous women to respond to the issue of non-consensually shared intimate images.[15]

- We supported Susan McLean and CyberSafety Solutions to deliver online education to students and parents across Australia. We supported continued education and resources for parents in a new online format, with greater capacity at the beginning of the pandemic.

- To support parents to understand the tools that are available on Instagram, we worked with ReachOut to develop an Instagram Parents Guide that contains suggested conversation starters to better understand how their teens are using Instagram and how to ensure they are using it safely and positively. We released the Guide in September 2019 and updated it in June 2021.[16]

[14] R Thomas, 'Young People at the Centre', Facebook Australia blog, 8 February 2021, https://australia.fb.com/post/young-people-at-the-centre/.
[15] Alannah & Madeline Foundation, *Helping Sistas be safer*, https://www.amf.org.au/news-events/latest-news/helping-sistas-be-safer/
[16] J Machin, 'A Parent's Guide to Instagram', *Facebook Australia blog*, 22 June 2021, https://australia.fb.com/post/a-parents-guide-to-instagram-in-partnership-with-reach-out/.

Encryption

Given the terms of reference specifically mention the impact of encryption on law enforcement, we provide some additional information about encryption and child safety here.

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, end-to-end encryption also poses a legitimate policy question: how to promote the safety of users if you're not able to see the content of their messages?

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."[17]

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 [emphasis added]****, 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.*
>
> *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*

[17] 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/.

Carnegie Endowment working group on encryption, polarized, absolutist positions in this debate should be rejected."[18]

The solution is for law enforcement and security agencies and industry, to work towards 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. This Committee has an opportunity to encourage a more nuanced debate in Australia about how to ensure the safety of users in an environment where virtually all communications are end-to-end encrypted.

Facebook has been continuing our industry leadership in combatting online CSAM by innovating and testing solutions that can detect possible online CSAM and take action, even if a service is end-to-end encrypted. Some of the new innovations outlined in the earlier section, just as limiting inappropriate interactions between adults and children, are agnostic about whether the messaging service is encrypted or not.

WhatsApp has been working in this space for some time and, even though it is end-to-end encrypted, we have been disabling more than 300,000 WhatsApp accounts per month for suspected sharing of online CSAM.[19] WhatsApp has also been submitting CyberTips to NCMEC. This detection is occurring via WhatsApp using advanced technology to proactively scan unencrypted information - including user reports - and to evaluate group information and behaviour for suspected sharing of CSAM.[20]

Facebook has indicated our intention to apply end-to-end encryption to Facebook Messenger; however, we know there is more to do to work through questions about how to continue our deep commitment to child safety on end-to-end encrypted services. When we announced these changes in early 2019, we publicly committed to a multi-year process of consultation to develop the most advanced safety mitigations possible for end-to-end encryption. That consultation process has involved consultation and engagement with Australian stakeholders, and it is continuing. Continued consultation with experts will help us bring our industry-leading track record on safety to an end-to-end encrypted environment.

[18] 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
[19] ASPI, *In-conversation with Will Cathcart*, https://www.youtube.com/watch?v=2KBQCsLDoBA.
[20] WhatsApp, 'How WhatsApp Helps Fight Child Exploitation', *WhatsApp Help Center*, https://faq.whatsapp.com/general/how-whatsapp-helps-fight-child-exploitation/?lang=en.