Australian Government | eSafetyCommissioner

# Submission

## Law enforcement capabilities in relation to child exploitation

**20 August 2021**

# Contents

# The eSafety Commissioner

The eSafety Commissioner (eSafety) is Australia's national independent regulator for online safety. Our core objective is to minimise harm to Australians online.

eSafety is the first government agency in the world dedicated specifically to online safety. We lead, coordinate, educate and advise on online safety issues and aim to empower all Australians to have safer, more positive online experiences.

When eSafety was formed in July 2015 (as the Children's eSafety Commissioner), one of the agency's main functions was administering a new regulatory scheme in relation to serious child cyberbullying. eSafety also assumed responsibility for the Online Content Scheme set out in Schedules 5 and 7 to the *Broadcasting Services Act 1992* (Cth), and previously administered by the Australian Communications and Media Authority.

Since then, eSafety's functions have broadened to include administration of a civil penalties regime in relation to image-based abuse ('IBA', sometimes referred to as 'revenge porn'), the power to issue notices to content and hosting services about abhorrent violent material, and a function related to blocking websites providing access to certain terrorist content during an online crisis event.

Beyond the protections built into our authorising legislation to provide take down of harmful content and deliver compassionate citizen service, prevention through awareness and education and initiatives to promote proactive and systemic change are fundamental elements to our successful regulatory model.

In drafting this submission, we have had regard to items (a) and (e) of the Inquiry's terms of reference, along with several related matters.

## eSafety's role in relation to online child sexual exploitation material

As Australia's online content regulator, eSafety plays a unique role within the Australian response to Internet-enabled child sexual exploitation. Our approach to the issue works across several axes.

### Online content reports and CSEM takedown[1]

We take public reports about online child sexual exploitation material (CSEM) and other harmful content for regulatory investigation and removal under the Online Content Scheme (explained further on page 5). Of the investigations we carry forward from these reports, 99% concern CSEM and all but a handful of these items are notified to the International Association of

[1] A note about terminology. Based on the ECPAT Terminology Guidelines (also known as the Luxembourg Guidelines), the term 'child sexual exploitation material' is a broad category of content that encompasses material that sexualises and is exploitative to the child, but that does not necessarily show the child's sexual abuse. Child sexual abuse material, which shows a sexual assault against a child, is a narrower category and can be considered a sub-set of CSEM. The eSafety Commissioner receives reports about material that is both sexually exploitative and that depicts child sexual abuse. For sake of simplicity, we shall refer to CSEM throughout this submission.

Internet Hotlines (INHOPE) network for rapid removal within the host jurisdiction.[2] This serves to alleviate harm to victims and survivors, who experience re-traumatisation as a result of the images of their abuse being circulated online. The Online Content Scheme also seeks to reduce the risk of end-users accessing or being exposed to this harmful content.

## Image-based abuse reports

Through the Image-based Abuse Scheme, we provide direct assistance to victims and survivors whose intimate images or videos have been shared (or threatened to be shared) without their consent. See page 6 for more information. About 25 – 30% of all IBA reports are made by Australians under the age of 18 years. Many of these reports appear to be linked with grooming and coercive behaviours. Removal is a key part of reducing the risk of ongoing harm to the children and young people who seek help from eSafety but there are cases where referral to relevant law enforcement agencies is warranted.

## Australian law enforcement agencies – memoranda of understanding

In late 2020, the eSafety Commissioner concluded a memorandum of understanding with the Australian Centre to Counter Child Exploitation (ACCCE). This is a crucial agreement for the eSafety Commissioner and establishes the Australian Federal Police (AFP) as eSafety's Commonwealth law enforcement partner. The MOU addresses how and under what circumstances eSafety will notify the ACCCE about threats to children. For example, where a matter reported to us as IBA appears to involve grooming, or where CSEM reported through the Online Content Scheme depicts an identifiable child or offender. In addition, the MOU establishes how the eSafety Commissioner works collaboratively with the ACCCE on prevention, education and communications that touch on areas of mutual concern.

In addition, we have MOUs in place with every state and territory police force. These MOUs deal with a variety of matters, including notification and referral of CSEM which concerns a specific jurisdiction. For example, if CSAM were to be hosted in New South Wales, eSafety would notify NSW Police prior to removal action. Once NSW Police was satisfied that operations or investigations would not be prejudiced by removal, eSafety would proceed with takedown. We are in discussion with several states to update and refresh these agreements in preparation for the Online Safety Act 2021 (see below).

[2] The International Association of Internet Hotlines (INHOPE) is a membership organisation consisting of 46 anti-CSEM hotlines around the world. Members include the US National Centre for Missing and Exploited Children (NCMEC), the UK's Internet Watch Foundation (IWF), and France's Point de Contact. INHOPE's vision is an Internet free from child sexual abuse material, and the association works closely with domestic, international and European law enforcement (including INTERPOL and EUROPOL) to share intelligence and contribute to victim identification efforts. INHOPE was formed in 1999, and the Australian Government has been a member (first through the Australian Broadcasting Authority, then the Australian Communications and Media Authority, now the eSafety Commissioner) since 2000. Members include industry associations, charities and public authorities (including the eSafety Commissioner and the Korean Communications Standards Commission).

## Prevention and education efforts

eSafety has a legislated role as the leader and coordinator of online safety education in Australia. This requires a comprehensive approach to producing guidance that addresses a range of online risks, for a variety of audiences.

Our statutory functions include supporting and encouraging measures to improve online safety for Australians; supporting, encouraging, conducting, accrediting and evaluating educational, promotional and community awareness programs relevant to online safety for Australians; and coordinating the activities Commonwealth Departments, authorities and agencies relating to online safety for children.

eSafety's education and prevention resources are evidence-based and provide extensive advice to children, young people, parents/carers and educators about a wide variety of online safety issues. We also have specialised resources for communities that may be marginalised or at greater risk of experiencing online harm.

The eSafety website includes advice about unwanted contact and grooming, how to report online exploitation (including to the AFP), and how to manage hard-to-have conversations with children about online safety. eSafety offers webinar-based training for teachers, parents and young people, including in the current series "Dealing with online harassment and image-based abuse", for parents, and "Online boundaries: it's ok to say no" for young people. This training has reached hundreds of thousands of parents, teachers and carers in the past year.

Drawing from our substantial in-house research, and collaboration with the education and early learning sector, we know that young children are increasingly given access to digital devices. By the age of four, 94% are already online. In response, eSafety provides a range of downloadable resources including a guide to online safety for parents and carers, and a set of Early Years materials. These support teaching online safety to children under five, while encouraging parents to stay engaged with their children's online lives.

As part of eSafety's role to coordinate and lift pedagogical standards in teaching online safety, we have recently published a *Best Practice Framework for Online Safety Education,* laying the foundation for a consistent national approach to education and prevention. The Framework identifies key pillars that should be in place for effective learning, including a strengths-based and age-appropriate curriculum, online safety principles taught at every year of schooling, and a balanced approach to risk and harm.

## Safety by Design

Finally, eSafety has spearheaded the global roll-out of the Safety by Design initiative. Safety by Design focuses on the ways technology companies can minimise online threats to users – especially younger users – by anticipating, detecting and eliminating online harms before they occur. Embedding safety into online products and services as core features from the very outset of product design is at the heart of the Safety by Design ethos.

Key to the initiative is a framework built around principles covering platform responsibility, user empowerment, and transparency and accountability. The principles have now been translated into a set of comprehensive tools allowing companies – from start-ups to established enterprises – to evaluate the current safety of their systems, processes and practices. The tools were developed with and for industry, highlighting industry best practice in innovations for safety.

Through Safety by Design, eSafety is seeking to lift the safety standards and practices of the technology industry to ensure greater protection of users and to minimise future threats. Safety by Design is intended to shift responsibility back to the platforms for safeguarding their users and engineering out misuse before harm occurs, rather than retrofitting fixes once the damage has been done.

### Regulating online harms

There are many departments and agencies at both the Commonwealth and state/territory level that share responsibility for combatting child exploitation and abuse. Important steps have been taken in Australia to create an integrated approach to tackling this harm, including where it occurs online. These steps include the watershed recommendations made through the Royal Commission into Institutional Responses to Child Sexual Abuse, the establishment of the National Office for Child Safety, and the creation of the AFP-led ACCCE.

Australian law enforcement agencies are at the very leading edge of global efforts to combat CSEM. National Joint Anti Child Exploitation Teams and specialists attached to the ACCCE work tirelessly to rescue victims and identify offenders. Over just two national operations – Operation Molto and Operation Arkstone – police arrested scores of Australians for child exploitation and laid hundreds of charges. Most importantly, at least 18 young victims were identified and made safe.

Police are to be commended for this difficult and critical work. However, law enforcement agencies cannot be expected to shoulder the effort of combatting CSEM alone. The flood of images and videos circulating on the Internet risks creating a permanent record of the abuse experienced by survivors – putting them in danger and exposing their suffering to the world at large.

As Australia's INHOPE hotline and online safety regulator, eSafety plays a complementary role to law enforcement in relation to taking down child sexual abuse imagery, while also providing direct support to young victims and survivors of image-based abuse through a civil scheme.

Many other hotlines within the global takedown network play similar roles. Public reports are encouraged through the ability to notify online CSEM anonymously, without the risk or fear of self-incrimination through a police-led reporting portal. Along with well-trained personnel, hotlines' strong and productive relationships with law enforcement support the effective management of risk. INHOPE hotlines and sister agencies contribute media and metadata to victim identification image libraries, including INTERPOL's International Child Sexual Exploitation Database (ICSE). In addition to eSafety, major global hotlines include the UK's Internet Watch Foundation (IWF), the US National Centre for Missing and Exploited Children (NCMEC), and the Canadian Centre for Child Protection (C3P).

We recognise that eSafety is part of a cross-agency, cross-sector, and multi-jurisdictional effort – one which has grown increasingly effective over recent years. To contribute to this effort, the eSafety Commissioner exercises a variety of regulatory powers.

## Online Content Scheme

Schedules 5 and 7 to the *Broadcasting Services Act 1992* (Cth) (BSA) establish the Online Content Scheme. Among other things, the Scheme provides eSafety with the power to regulate the hosting of prohibited content in Australia. Whether content is prohibited is a decision made with reference to the National Classification Scheme applicable to films. Material hosted in Australia that is classified Refused Classification (RC) or X18+ will be prohibited, while material classified R18+ will be prohibited unless it is subject to a restricted access system.

Prohibited content is subject to a takedown notice, issued by the eSafety Commissioner. Takedown notices are issued against the relevant Australian hosting service provider, and must be complied with by 6pm the following business day. Non-compliance attracts a civil penalty.

As a result of the strong civil regulatory and criminal enforcement framework in Australia, prohibited material – including CSEM – is rarely hosted here. Accordingly, since 2015, the eSafety Commissioner has issued only a single takedown notice in relation to Australian-hosted prohibited material, where R18+ material was provided via an Australian-hosted adult website. Overwhelmingly, CSEM is hosted overseas and predominantly within INHOPE member jurisdictions.

Under Schedule 5 to the BSA, the eSafety Commissioner must notify Australian law enforcement in relation to overseas-hosted 'sufficiently serious material' (such as CSEM). However, so long as there is an agreement in place with an Australian police commissioner, the eSafety Commissioner may notify such material to another person or body. Through the eSafety/ACCCE MOU, eSafety has secured agreement that CSEM hosted in a country within the INHOPE Network is notified to INHOPE, with URLs hosted in other countries reported to the AFP on a regular basis. This continues a long-standing practice agreed to with the AFP since the Australian Government joined INHOPE in 2000.

In the financial year 2020/21, eSafety notified almost 13,000 CSEM items to INHOPE for removal and law enforcement action in the host jurisdiction. Media and metadata relating to verified CSEM reports processed by INHOPE are shared with INTERPOL for inclusion in its victim identification database, ICSE.

## Image-based Abuse Scheme

Part 5A of the *Enhancing Online Safety Act 199* (Cth) (EOSA) sets out a regulatory scheme for investigating and acting against complaints about the non-consensual distribution of intimate images. Section 9B of the EOSA defines an intimate image as including where the image depicts or appears to depict a person's genital or anal area (including when covered by underwear), or a person's breasts if the person identifies as female, transgender or intersex, in circumstances in which an ordinary reasonable person would reasonably expect to be afforded privacy. Material is also an intimate image if it depicts a person in certain forms of private activity (for example, in a state of undress, using the toilet or showering) in private circumstances. In cases where a person's cultural or religious background involves the wearing of certain religious attire, an image will be intimate if it shows that person without the attire in a private setting.

There will be a contravention of the EOSA when a person posts or threatens to post intimate material without consent. Under the EOSA, consent to share intimate material cannot be given by a child under the age of 18. To be captured within the IBA scheme, material must be posted on (or the threat must relate to) a social media service (such as Facebook), a relevant electronic service (including messaging services such as WhatsApp), or a designated Internet service (which includes websites) and either the perpetrator or victim (or both) must ordinarily reside in Australia.

eSafety has a number of regulatory options in relation to IBA which can be directed at either the service providing access to the material or the person responsible for posting (or threatening to post) it. In cases involving a child victim and a perpetrator who is or may be an adult, eSafety is more likely to notify the perpetrator to law enforcement than to take civil action against them. The way we respond to these cases is explained in more detail below.

## The Online Safety Act 2021

A major reform to the regulation of online harms will commence in January 2022 through the *Online Safety Act 2021* (Cth) ('OSA'). The OSA is intended to create a modern, fit for purpose regulatory framework that builds on the existing legislative schemes for online safety. Relevantly the OSA:

- strengthens the existing Online Content Scheme by expanding the number of services relevant to its operation, and providing the eSafety Commissioner the power to issue removal notices against 'class 1' content (which includes CSEM) wherever that content is hosted, globally
- creates new powers for the eSafety Commissioner to direct online app stores and providers of online search services to remove apps and delete links that allow access to that material where one or more class 1 removal notices have been ignored

- introduces a set of Basic Online Safety Expectations through a ministerial legislative instrument that will allow the eSafety Commissioner to require transparency reporting on how services are keeping their users safe, including how they are preventing their platform from being used as a vehicle for CSEM
- provides for the creation of one or more industry codes or standards to promote the adoption of responsible industry processes and procedures for dealing with online content issues, including CSEM.

While the provisions that relate to IBA are substantially similar to those set out in the EOSA, the interval for a service to respond to a removal notice will be reduced from 48 to 24 hours – a feature now applicable across all the OSA schemes. In addition, the OSA creates a world-first scheme to address seriously harmful adult cyber abuse, an enhanced cyberbullying scheme for Australian children and young people, and improved information-gathering powers. eSafety has produced a fact sheet on the OSA, available here.

## The problem of child sexual exploitation material

The phenomenon of producing and sharing child sexual exploitation material pre-dates the Internet. However, the pre-online trade came with significant risks to offenders, reliant as it was on distributing hard copy material either through the post or via small interpersonal networks. Processing photographs and film depicting the sexual abuse of children presented considerable risk, given the need to outsource to film processing labs. In consequence, the demand for material through this period was frequently catered to by child sexual exploitation magazines with names such as *Lolita* and *Nudist Moppets.*

With the advent of dial-up Internet, the opportunity to connect with likeminded offenders with relative ease and anonymity increased substantially. Digitised versions of CSEM imagery, often scanned from magazines, were shared on bulletin boards and via email. However, file sizes were still limited by dial-up connection speeds and shaky infrastructure.

Connection speeds and bandwidth improved through the early 2000s. Alongside this technical development, digital cameras became affordable household items. It did not take long before digital cameras were integrated into mobile phones and, later, smartphones. The Internet began to abound with images produced and shared by offenders abusing children in their care. Websites, peer-to-peer networks, imageboards and forums became common and highly accessible locations to encounter CSEM.

The scale and scope of child sexual exploitation online is staggering. Far from being a threat that exists solely on the ‘dark web’, this is all too often a crime and form of abuse that is playing out in front of us. The ‘clearweb’ (that part of the Internet that is indexed and can be reached by common browsers) remains a preferred medium for the distribution and hosting of CSEM at scale. On the clearweb, well-known top-level domains such as .com and .net are routinely abused to host CSEM, and open websites provide access to hundreds of thousands of images.

The figures speak for themselves. In 2020, our sister hotline in the UK, the IWF acted on close to 155,000 reports of child sexual abuse imagery. Almost half of these reports related to ‘self-generated’ imagery (including children recording themselves performing sexual acts) – an increase of 77% on 2019. The IWF explains that some of these images appear to have been

created within the context of a romantic relationship between peers, but later shared more widely online. Other images show evidence of being created through coercive, manipulative and exploitative interactions with adults.

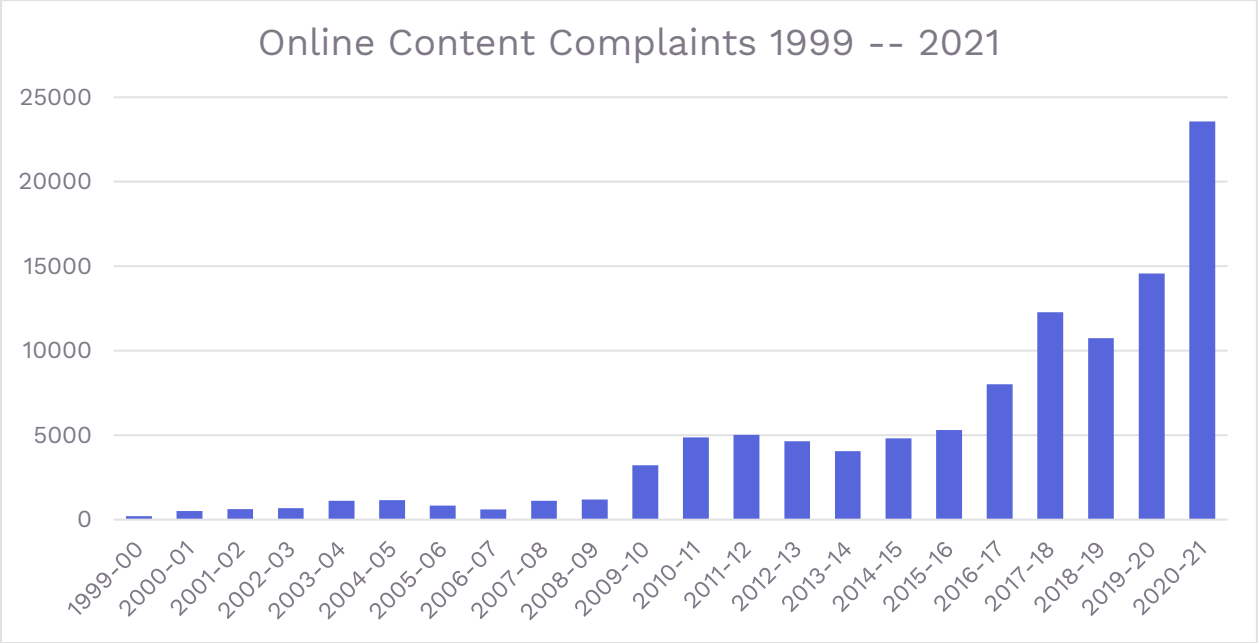
The Canadian Centre for Child Protection (C3P) has, through its Arachnid program,[3] detected and verified more than 5.4 million child sexual abuse images since 2018. Through the program, C3P has notified more than 760 electronic service providers worldwide that they are hosting CSEM. Almost 85% of the images identified through the program relate to victims that are not known to have been identified by law enforcement agencies. eSafety has partnered with C3P and contributes to the work of the Arachnid program through classification and verification of detected CSEM images, helping accelerate Arachnid's automated removal of CSEM at-scale.

During 2020, the INHOPE network exchanged reports about more than one million URLs depicting suspected CSEM. More than 90 percent of the content showed the abuse and exploitation of girls, and just over three quarters of all reported CSEM involved the abuse or exploitation of pre-pubescent children. Almost all content reported as being provided from Europe was hosted in the Netherlands.

## Complaints about CSEM made to the eSafety Commissioner

Over the more than 20 years of the Online Content Scheme's operation, complaints about illegal and offensive content by the public have seen a steady increase. During the first full year of the Scheme's operation, 201 public reports were received about a variety of content types. In financial year 2020-2021, the eSafety Commissioner received more than 23,500 public reports about offensive and illegal online content. This was an increase of more than 60% on the previous financial year. Overwhelmingly, public reports concern child sexual abuse material.

[3] The Arachnid program crawls the World Wide Web for known child sexual abuse material (and related imagery) enabling automated removal notices to be sent to providers. The eSafety Commissioner participates in the Arachnid program, assisting with the classification of images identified during crawling. Arachnid is a collaboration between C3P, the Royal Canadian Mounted Police, and participating hotlines. More information can be found at https://projectarachnid.ca/en/.




Over time, eSafety has observed a distinct shift in the nature of CSEM identified through regulatory investigations, and the nature of hosting by industry. Images and videos are far more likely to have been produced by children and young people themselves, often involving explicit sexual posing and sexual touching. This type of content appears in substantial volumes on websites and forums catering to those with a sexual interest in children, and appears to often have been produced through trick, threat or manipulation.

Increasingly, CSEM websites are hosted by network providers that deliberately obscure their corporate footprint. This obfuscation can be achieved by providers registering company details in jurisdictions such as the Seychelles, distributing registration across jurisdictions, and deliberately undermining the integrity of the global WHOIS database. Some providers openly market themselves as being 'bulletproof': resistant to takedown and disruption and with a high tolerance to hosting illegal content. Removal of CSEM sites by INHOPE members, industry and law enforcement can be complicated by these tactics.

## Classification of material on streaming services

The Australian Classification Board has worked with Netflix to create a tool allowing classification of Netflix content that is compatible with the National Classification Scheme. A 2018 review of the tool found that it produced decisions that were broadly consistent with the National Classification Scheme in 93% of cases.[4] The classification of material across delivery formats (including streaming services) will be considered by the review of Australian classification regulation currently being undertaken by the Department of Infrastructure, Transport, Regional Development and Communications.

[4] Commonwealth Department of Communications and the Arts, 'Report on the Pilot of the Netflix Classification Tool', <https://www.classification.gov.au/sites/default/files/2019-11/report-on-pilot-of-netflix-classification-tool_0.pdf>, 4.

eSafety has not encountered a significant problem with the classification of material on commercial streaming services such as Stan, Netflix, or Foxtel Now/Binge. During financial year 2020-21, eSafety received 2 complaints about material available on the Stan service, however the material was not deemed sufficiently serious to warrant an investigation. In the same period, we received 30 complaints about Netflix. Most of these complaints concerned *Cuties*, a film by French director Maïmouna Doucouré about an eleven-year-old Senegalese-French girl.

The film deals with various themes, in particular the hyper-sexualisation of pre-adolescent girls. While the film attracted considerable controversy for its depiction of this theme, the Australian Classification Board and Netflix tool classified the film MA15+ (Mature Accompanied). The rating's consumer advice included a warning about 'Strong themes'. Based on this rating, eSafety did not judge *Cuties* to be sufficiently serious to warrant an investigation.

## Image-based abuse complaints

eSafety is the only regulator in the world to oversee a legislated civil penalties scheme for image-based abuse. Reports to eSafety about image-based abuse have also risen since the commencement of the civil penalties scheme in September 2018. About 25-30% of reports about IBA are made by those aged under 18 years. Most under-18 reporters are aged between 13 and 17 years, with only a small percentage (7%) under 13.

Of the reports received from under 18s, most concern online child sexual exploitation. Only 8% concern peer-group sharing. Young reporters are typically coerced into sharing images of themselves by adult offenders, who are often pretending to be young people. Once a young person has sent an image to this type of offender, threats to share their images are received and demands are made for further images. We have developed procedures which ensure eSafety is a safe place for children and young people to come for help with these matters. These procedures align with our obligations to provide relevant information to police, including to the ACCCE.

eSafety is strongly committed to working with police to hold offenders accountable and we regularly notify information to achieve this shared objective. We manage risks to the relevant child or young person by ensuring that they cease all contact with the offender, and we work with the relevant online platform to have the child's image and/or the offender's account removed (in consultation with the ACCCE, where relevant). Over the life of the IBA scheme, eSafety has alerted social media services to the misuse of almost 500 accounts involved in the sexual exploitation of a child or young person, with services disabling over 80% of the accounts reported. We also refer children and young people to Kids Helpline for counselling and support.

Where peer-group sharing is relevant to a report, we have found that a law enforcement approach is not always a preferable option for resolution. While these matters are typically reported to police by either school staff or parents, police for a number of reasons do not always elect to prefer charges. This decision might be due to insufficiency of evidence, or the age and vulnerabilities of the children involved. We typically address this type of matter by:

- reporting accounts that have shared, or threatened to share, intimate images to the social media service
- giving advice on how the victim can screenshot evidence (for example, of threats or account profiles) and block accounts
- providing safety advice regarding privacy settings and deleting all friends/followers who are not known and trusted offline.

We might also:

- liaise with schools if they are in a position to help address the incident
- speak with police if they are already involved or ought to be involved

- take remedial action, for example, by writing to the young person/s responsible for the non-consensual sharing, warning them that their actions are unlawful and requiring confirmation that they have deleted the intimate images from their devices and anywhere they may have posted them online.

eSafety has received more than 6,400 reports about IBA over the life of the scheme.

Almost 70% of all reports have been received in the last 18 months alone.


eSafety's research shows that Australian teens are exposed to a range of risks and threats online. More than 40% of young Internet users report negative experiences online. These include being contacted by a stranger (30%) and receiving inappropriate or unwanted content such as pornography (20%).[5] While many teens take some form of action against the unwelcome contact, less than half mention it to family or friends (43%) or report it (40%).[6] Online safety information is valued by teens, with three-quarters wanting information about issues such as how to block bad actors, how to support friends in trouble, and how to report negative online experiences.[7]

All of this makes clear that the prevalence and accessibility of CSEM online is a challenge that goes well beyond law enforcement. Instead, addressing the many elements that enable the online sexual exploitation of children demands a whole-of-government, whole-of-community

[5] eSafety Commissioner, *The digital lives of Aussie teens*, <https://www.esafety.gov.au/sites/default/files/2021-02/The%20digital%20lives%20of%20Aussie%20teens.pdf>, 5.
[6] Ibid, 6.
[7] Ibid.

approach that reaches across borders and jurisdictional limits. The eSafety Commissioner plays an active role in this response through our regulatory interventions, education and prevention initiatives, and policy innovations such as Safety by Design.

# The role of technology providers in assisting law enforcement and governments

## Industry's policies overall

Most mainstream services have policies, rules, terms of use or community standards prohibiting child sexual exploitation and abuse on their platforms. When they become aware of such content, mainstream services which are subject to US federal law typically remove it, disable the relevant account, and report it to NCMEC. NCMEC forwards the reports to law enforcement agencies around the world, including the AFP. In 2020, NCMEC received 21.4 million reports from electronic service providers related to suspected child sexual exploitation shared via their networks or held in their data storage systems.[8]

Services detect and action CSEM in a variety of ways, including through Trust and Safety teams and automated tools. Some of this work is proactive, such as scanning content for potential CSEM at upload, and some is reactive, such as providing reporting mechanisms for users to notify potential CSEM to the service. As outlined below, the effectiveness of these measures varies across services, as does the level of investment, innovation and collaboration undertaken to combat CSEM.

Another variable element is the level of transparency that services provide in relation to these efforts. Many transparency reports remain centred on government requests for content removal. However, services are increasingly beginning to report on the amount of CSEM discovered on their platforms through proactive tools and user reports, in addition to the items surfaced through government notices. Reports may also set out the number of accounts disabled and items of content removed and reported to NCMEC, as well as providing details about other initiatives, projects and partnerships in this space.

There are several groups currently working to drive up industry practices and standards through collective action. These include the industry-led Technology Coalition and the cross-sector, multi-stakeholder WePROTECT Global Alliance (WPGA). The eSafety Commissioner serves as a member of the WPGA Board and recently coordinated Australia's response to the WPGA's survey on implementation of the Model National Response, a blueprint for national action to tackle online child sexual exploitation.

[8] National Centre for Missing and Exploited Children. '2020 Reports by Electronic Service Providers (ESP), <https://www.missingkids.org/content/dam/missingkids/gethelp/2020-reports-by-esp.pdf>.

## eSafety's experience in working with industry on CSEM issues

Efforts by major industry representatives to harden their platforms and networks are welcomed by eSafety. Several initiatives – some longstanding – have had a tangible impact on the ability of offenders to find, share and store CSEM online.

They include:

- Google: In many countries, users who attempt to locate CSEM via Search are met with Google Ads showing deterrence messaging. In Australia, this messaging warns users that the 'intentional viewing or possession of sexually explicit imagery of minors is illegal'. A reporting link to the eSafety Commissioner is provided, along with information about contacting the AFP and assisting victims of child sexual abuse through Bravehearts. Google also provides its Content Safety API – an artificial intelligence classifier for CSEM – to customers for free. The API is intended to help organisations scale and prioritise decisions around content remove content. YouTube also freely offers its CSAI (Child Sexual Abuse Imagery) Match technology, allowing for detection and matching of known child sexual abuse video content.
- PhotoDNA: A key tool in the identification and removal of CSEM at scale is PhotoDNA. This is a 'hashing' technology able to convert images into a unique signature. This signature can be used to find similar images, and is used widely by industry and NGOs such as C3P and NCMEC to detect, notify and remove known CSEM. PhotoDNA was developed in 2009 through a partnership between Microsoft and Dartmouth College in the United States. The technology is offered free as a cloud service to qualified organisations.
- Other hashing technologies: Facebook has released two hashing algorithms as open-source projects to assist with detecting CSEM. Known as PDQ and TMK+PDQF, the algorithms allow for perceptual hashing of images and videos, respectively. Both are offered free from a public GitHub repository.
- Project Artemis: An anti-grooming tool developed by Microsoft in collaboration with The Meet Group, Roblox, Kik and Thorn. The tool is made freely available by Thorn to qualified organisations that offer a chat function as part of their service. Artemis helps with moderation of high-risk conversations on platforms that flag potential grooming efforts, and is based on technology originally deployed by Microsoft on its Xbox gaming platform.
- Apple: Recently, Apple announced its commitment to preventing its products and services from being misused to distribute CSEM. Starting with efforts to limit the potential for children to come to harm using Apple technology, the company will soon add new tools to warn children and their parents when receiving or sending messages containing nudity. In addition, on-device hashing of images will now occur in a way that preserves privacy while allowing detection of CSEM. Finally, Apple will provide warnings and information to those who attempt to search for CSEM using Apple services.

However, there are still areas that warrant improvement.

For example, in early 2021 the Canadian Centre for Child Protection (C3P) analysed the reporting functions provided to users by major platforms.[9] While most platforms provided a way for users to report illegal or inappropriate content, there were few cases where a CSEM-specific option was provided. In addition, C3P identified several features that created inhibitions against reporting, such as requiring users to provide personal contact information, requiring users to create an account before being able to flag content, and an inability to report specific users, profiles, posts or a combination.

In 2020, eSafety identified a number of accounts on a major platform that appeared to have been created for the sole purpose of sharing CSEM. The accounts were often private but displayed specific indicators that strongly suggested their purpose. For example, many referred to popular file-hosting platforms such as Mega, displayed images of known CSEM victims in their profile, and contained text such as 'DM to trade' and 'cheese pizza' (shorthand for 'child pornography'). Even though no content was posted to these accounts, they often had follower counts in the thousands. At the time, eSafety noted that there was no way to report entire accounts for CSEM-related violations.

Shortly after discussing its internal report with the ACCCE, eSafety sought a meeting with senior platform representatives. During the meeting, the eSafety Commissioner explained the key indicators we identified as suggesting that accounts were CSEM-related and explained our concerns with the sufficiency of reporting options. The company representatives undertook to review their processes and some changes were made to detection and reporting procedures. We have observed a reduction – but not an elimination – of these kinds of reports.

eSafety remains concerned at the lack of progress made within industry overall on the issue of content that is related to but does not depict CSEM. Overwhelmingly, survivors of online child sexual abuse are concerned about the potential for their abuse material to become known to those in their lives. More acutely, many survivors fear recognition by strangers from their abuse material. Sadly, this is all-too-often a fear that is justified, with 30% of survivors surveyed in a 2017 study by C3P disclosing that they had been identified online or in-person by someone who had seen their abuse imagery.[10] Survivors have been physically followed, threatened and propositioned as a result of being recognised and targeted.

While industry tends as a rule to remove clear CSEM from its networks and storage services, there is far less commitment to removing related material. The sexual abuse and exploitation of children online frequently occurs within a context of an image series showing the child dressed, and then in various states of undress prior to the depiction of contact offending (for example penetrative sexual assault). The 'scene-setting' images within a series can be just as harmful to survivors when available online, as they form part of a continuum of abuse that remains fresh and distressing. Even though they may not be illegal per-se, the images are a reminder of trauma and warrant removal.

However, it can be a challenge for hotlines and others working in content removal from a victim perspective to persuade industry that these images should be removed. Often, industry will remove material only when it is illegal within a specific jurisdiction, and in some cases efforts to

[9] Canadian Centre for Child Protection, *Reviewing Child Sexual Abuse Material Reporting Functions on Popular Platforms*, <https://protectchildren.ca/pdfs/C3P_ReviewingCSAMMaterialReporting_en.pdf>, 8.
[10] Canadian Centre for Child Protection, *Survivors' Survey Full Report,* <https://protectchildren.ca/pdfs/C3P_SurvivorsSurveyFullReport2017.pdf>, 165.

take down CSEM-related material are met with resistance. There is also reluctance to removing written accounts of adults sexually abusing children or illustrated and drawn depictions of sexual abuse (even though they are prohibited in several jurisdictions including Australia). We note that internationally a schism is forming around content that is 'illegal' and content that still extremely harmful but is legal. Proposed legislation and regulatory approaches in the UK (Online Safety Bill), Canada (Discussion guide), Ireland (Online Safety and Media Regulation Bill) and the EU (Digital Services Act) grapple with this issue, to varying degrees.

We are concerned with using illegality as the vector to determine whether industry should act in response to harmful content. With this type of approach, a huge spectrum of online harms would fall through the cracks of regulation and response, ultimately leading to individual harm. Online platforms should retain the prerogative to identify harmful content based on users' complaints for illegal and harmful content, to safeguard children and all citizens online.

It can be seen, then, that there is still much work to do. Noting this, it is worth emphasising how critical a partner industry is in counter-CSEM efforts. The modern Internet – its wires, hardware, data centres, and cabling – is almost entirely owned and operated by private concerns. That means that efforts to harden the online world against abuse by those producing and distributing CSEM will only be effective with sustained and systemic buy-in from the network operators, domain registrars, Internet address registries, domain administrators, hosting service providers, enterprise cloud providers and others. This requires sustained cross-jurisdictional efforts and consistency of regulation, globally.

# Key Challenges

## Encryption

Digital encryption is not new and, in its modern form, has been used for more than 40 years as an essential tool for privacy and security. It is primarily employed to keep data and transactions secure and to prevent data breaches and hacking. It allows legitimate, positive and safe communication where this may not otherwise be possible, and is used to protect valuable information such as passport credentials.

However, encryption can also assist in serious harms by hiding or enabling criminal activities, including online child sexual abuse. Technologies that detect illegal material by proactively scanning, monitoring and filtering user content currently do not work on systems that use end-to-end encryption (E2EE). Because of this, E2EE can facilitate the production, exchange and proliferation of child sexual abuse material, perpetuating the abuse of victims and exposing survivors to ongoing trauma.

A drift towards E2EE by major social media platforms will make investigations into serious online child sexual abuse and exploitation significantly more difficult. It will create digital hiding places, and platforms may claim they are absolved of responsibility for safety because they cannot act on what they cannot see.

We know there are a number of solutions that would ensure illegal activity online can be addressed. These work without compromising encryption while allowing lawful access to information needed in serious criminal investigations. Solutions include using certain types of encryption that allow proactive tools to function, implementing proactive detection tools at transmission, rather than on receipt, and moving AI and proactive technical tools to the device level (as Apple is doing).

## Anonymity and identity shielding

Anonymity and identity shielding allow a user to hide or disguise their identifying information online. Anonymous communication is a cornerstone of promoting freedom of speech, expression and privacy on the Internet, but it can also be misused to control and abuse people.

Technical approaches to anonymity include software, browsers and encrypted or decentralised platforms. Examples include virtual private networks that mask the user's location and device details (IP address), anonymising processes that conceal the link between a message and the sender, and E2EE that allows only a sender and recipient to decode digital content.

Simpler approaches involve taking on a fictional identity. Examples include using a false name (i.e., a pseudonym or alias), a virtual representation (or avatar), or a fake profile.

Most investigations into CSEM involve individuals posting the content online anonymously. These investigations have shown that content contributors will go to great lengths to remain anonymous, often using one or more anonymising security measure to hide their identities.

Sexual predators also commonly use anonymous, fake, imposter and impersonator accounts to lure victims and gain their trust. For example, they may use an avatar in a game to pretend they are the same age and gender as a child so they can become a fake friend and groom them for sexual interaction.

It is very difficult for regulators and law enforcement to identify and act against individuals and using fake accounts. It also makes it almost impossible for social media services and other users to deal with abusers breaching the terms of service, through strategies such as blocking and suspension, as well as preventing, detecting and removing multiple accounts operated by one user.

A balance is needed, where the misuse of anonymity and identity shielding is restricted without removing any of the legitimate benefits. Steps can be taken by services to verify accounts before users start to operate them, or to take down accounts that violate the terms of service and prevent them from resurfacing.

## Decentralisation

Decentralisation of the Internet means widely distributing the control of the online data, information, interactions and experiences of users so they are no longer reliant on a concentration of large technology companies that own or operate mainstream, centralised servers (the computer hardware and software that stores data) to access the online world.

While decentralisation can allow users to protect their information and control their online experiences, it can also make it more difficult to hold users (or the entities behind them) responsible for illegal and harmful content and conduct. The lack of a central authority, along with the storage and distribution of data across many computers, makes it difficult to moderate or regulate decentralised services and platforms or enforce the removal of illegal and harmful content. For these reasons, there are concerns that a decentralised Internet may become a haven for CSEM and for users who have been removed from mainstream services and platforms.

As interest grows in the tech community to develop the 'DWeb' and 'DApps', and as mainstream platforms increasingly respond to and address CSEM on their services, the perceived impenetrability and unaccountability of decentralised environments could act as an incentive for those with nefarious intent to evade detection, to preserve their 'collections' of materials and to further create and distribute CSEM.

We must work collectively and across borders to encourage greater consistency and shared approaches to help counter online risks and harms on decentralised services and platforms. There is also need to ensure that safety-by-design is given the same priority as security- and

privacy-by-design in the design and development of decentralised services and in the broader Web 3.0 infrastructure.

There are a number of ways decentralised services and platforms can help to keep their users safe from online harms. For example, online communities can opt-in to moderation and governance arrangements. Features such as voting systems can allow users to decide acceptable conduct and accessible content. Additionally, built-in incentives, such as micropayments or other rewards, may encourage positive behaviour and safer environments. Decentralised services and platforms can also be built using technology protocols that allow third party content moderation tools to function. For example, tools that scan for child sexual abuse material might be adopted, though their operation would have to be agreed to by the community of users.

## Addressing Challenges through Safety by Design

eSafety recognises that encryption, anonymity and decentralised systems may help to protect certain elements of privacy and security. Our focus is on working with industry and developers to ensure that services are aware of Safety by Design principles and adopt them, so the risks of these features are considered along with the benefits.

The initiative has been developed with industry for industry. It recognises that, if we wish to end child sexual exploitation and abuse, industry needs to be at the heart of any process to effect cultural change through enhanced corporate social responsibility. eSafety has undertaken extensive consultation with industry, civil society organisations, advocates, parents and young people themselves to understand how online harms develop and are experienced across broad and intersectional communities.

As noted above, our Safety by Design principles have now been translated into a set of comprehensive tools allowing companies – from start-ups to established enterprises – to evaluate the safety of their systems, processes and practices. This includes advising industry on how to ensure that robust moderation of conduct and content is possible before releasing products to the market, as well as how to authenticate users and prevent known techniques used by perpetrators to target and abuse others.

Safety by Design encourages technology companies, and indeed the broader technology industry, to help end child sexual exploitation and abuse by enhancing their corporate social responsibility. In part, this can be done by highlighting the innovation that is already occurring within the sector as well as encouraging technology companies to foster a global community and to be open in sharing their solutions.

User-centred design with consideration of children and young people is critical. Key touchpoints for industry consideration include implementing default privacy and safety settings at the highest possible levels, incorporating conversation controls and discoverable and seamless reporting pathways. Such measures proactively address the potential for online harm, while empowering users to regulate their own online experiences.

eSafety continues to work closely with industry to further implement existing safety measures, standards, requirements and guidance – as well as encourage them to innovate and transform the safety landscape further. Our forward workplan for Safety by Design includes working with the investment community to incorporate the principles into responsible investment practices; generating practical engagement with the assessment tools within the start-up community; focusing on marginalised and at-risk groups to ensure their needs are considered; and developing targeted resources for new and emerging sectors.