Australian Government | eSafetyCommissioner

**03 September 2025** **CC25-0112**

Committee Chair
Joint Committee on Law Enforcement
Department of the Senate
PO Box 6100
Parliament House
CANBERRA ACT 2600

*Via email: le.committee@aph.gov.au*

Dear Committee Chair

The eSafety Commissioner (eSafety) welcomes the opportunity to update evidence provided previously to the Joint Committee on Law Enforcement's inquiry into the capability of law enforcement to respond to cybercrime, including cyber-enabled crime. The landscape of cybercrime evolves rapidly, and there have been significant developments since our initial submission of 15 December 2023, my evidence to the committee at the hearing on 22 October 2024 and our response to subsequent Questions on Notice.

In this update, we continue to concentrate on the following terms of reference:

a) Coordination efforts across law enforcement, non-government and private sector organisations to respond to the conduct of cybercrimes and risks of cybercrime
b) Emerging cybercrime threats and challenges affecting Australian entities and individuals, including the scale and scope of cybercrimes conducted in Australia or against Australians
c) Existing law enforcement capabilities in the detection, investigation and prosecution of cybercrime, including cyber-dependent crimes and cyber-enabled crimes
d) Prevention and education approaches and strategies to reduce the prevalence of victimisation through cybercrime, and
e) Other related matters.

**Information sharing**

The dynamic nature of online harm means that strategies and priority setting must be constantly reviewed to ensure we are addressing emerging issues. Information sharing is vital to raising awareness of serious online harm, harm trends and threats.

To share information and intelligence efficiently and effectively, Memorandum of Understandings and Letters of Exchange are in place with the Australian Federal Police (including the Australian Centre to Counter Child Exploitation (ACCCE)), and Australian police forces, setting out our respective areas of operation and processes for information sharing and cooperation.

*National Youth Crime Online Roundtable*

In June 2024, eSafety convened a National Youth Crime Online Roundtable to establish a shared understanding among law enforcement, industry, academia, advocates, judicial and government of the problem of children and young people inciting/glorifying crime and violence.

A key action item from the roundtable was the establishment of a Joint Task Group to be chaired by eSafety. In 2025, eSafety convened the first Joint Task Group, focusing on testing and evaluating the coordination and response capabilities of eSafety and law enforcement agencies in managing this type of material posted online. The initiative also played a key role in brokering partnerships between the technology platforms we work with daily and state police forces, strengthening collaborative efforts to tackle these crimes more effectively.

**Illegal and Restricted Content reports**

eSafety investigates complaints from the public about child sexual exploitation material (CSEM), pro-terror content and other forms of illegal or harmful online content. In FY 2024-25, eSafety received complaints about 47,292 URLs, representing an increase of approximately 30% on the number received in FY 2023-24. About 85% of all reports concerned CSEM.

Beyond the rising volume, the nature of the content is becoming increasingly complex and confronting. Reports have included highly deviant and sadistic forms of abuse, extreme violence, and disturbing imagery. While synthetic CSAM remains a small proportion of complaints, it is difficult to detect and expected to grow with the increasing availability of generative AI tools.

eSafety's intelligence and compliance teams have observed a broader spectrum of harms, including:

- The emergence of new platforms and services being used to share illegal content, often with enhanced features that facilitate rapid distribution and user engagement.
- A resurgence of forums and networks dedicated to non-consensual image sharing, some of which now incorporate AI-generated material.
- Graphic and violent content that, while not always actionable under current thresholds, raises serious concerns about desensitisation and online safety norms.
- Increasing crossover between online harms and other criminal domains, such as financial crime and extremist content, requiring deeper collaboration with law enforcement and intelligence partners.

These trends underscore the need for continued vigilance, adaptive enforcement strategies, and strong partnerships across government, industry, and civil society to address the evolving threat landscape.

**Terrorist and violent extremist content and online radicalisation**

The spread of terrorist and violent extremist content (TVEC) and its role in online radicalisation, particularly of young people, is a concern both in Australia and overseas.

Ever since the attack in Christchurch, eSafety has been particularly concerned about the role of livestreaming, recommender systems and now Artificial Intelligence (AI), in producing, promoting and spreading TVEC and activity.

eSafety has key powers under the *Online Safety Act 2021* (Cth) (OSA) to minimise some of these risks

and ensure online services fulfill their responsibilities. In particular, we now have codes and standards in force that create mandatory obligations on a wide range of services to proactively detect and minimise pro-terror material on their services.

In addition to the protections offered by the codes and standards, discussed in more detail below, the Government's social media minimum age legislation will put in place a critical delay for children under the age of 16 from having accounts on certain social media platforms, which aims to reduce their exposure to harmful content which is driven and amplified by opaque algorithms and deceptive design features.

*Growing risks posed by AI*

eSafety is concerned by reports that terrorists and violent extremists are moving to capitalise on the emergence of generative AI and are experimenting with ways this new technology can be misused to cause harm. The companies that provide these services have a responsibility to ensure that these features and their services cannot be exploited to perpetrate such harm.

eSafety recently used its world-leading transparency powers to compel some of the platforms where there are risks of this kind of material to answer questions about how they are tackling it. We released this information in March 2025 in a transparency report. Featuring Google (which includes YouTube and Gemini), Meta, WhatsApp, Telegram and Reddit, the report lays bare industry failures in addressing this issue.

*Operation Catalyst*

Operation Catalyst is an eSafety operation focused on investigating and removing targeted publications which have been, or likely would be, Refused Classification under section 9A of the *Classification (Publications, Films and Computer Games) Act 1995* (Cth) and would therefore be class 1 material under the OSA.

The targeted publications are written declarations likely to facilitate terrorist acts and other acts in preparation for, or in the planning of terrorist acts; advocating the doing of terrorist acts likely to cause serious physical harm to a person or property; or likely to cause a person's death or endanger a person's life or creates a serious risk to the health and safety of the public.

The operation specifically targeted 13 manifestos, many of which are the most well-known manifestos written by the perpetrators of terrorist attacks around the world and used as templates for terrorists since. Most of the material was located by searching the respective titles of the manifestos or through the description of the terrorist attacks with which they are associated.

Over a two-week period 174 copies of targeted manifestos were found across a broad range of online services including social media, file hosting and document sharing services as well as webpage archives.

To date, 137 copies of manifestos have been removed through content notifications, removal notices or the site being down. There are a further 38 copies remaining available with 13 of those awaiting classification by the Classification Board.

Next steps in Operation Catalyst include further removal notices, and possible link deletion notices and service provider notifications. eSafety will also use the information gathered to inform our systemic enforcement work as well as sharing an information report with external stakeholders.

**Image-based abuse (IBA) reports and sexual extortion**

eSafety's Image-based Abuse Scheme investigates and provides direct assistance to individuals whose intimate images or videos have been shared online (or threatened to be shared) without their consent. Most reports that eSafety receives under this scheme involve sexual extortion (sometimes called 'sextortion'), a form of financial blackmail where someone threatens to share a nude or sexual image or video of a victim unless they pay the perpetrator.

The general modus operandi for sextortion complaints reported to eSafety follows a sequence of strategically manipulative steps, as outlined below:

1. The perpetrator engages with and communicates with the victim via social media, dating or chat platform/s.
2. The perpetrator often:
    - Uses online platforms that provide the ability to identify a potential victim and harvest sufficient information about them, i.e. personal information, contact lists, and details about social networks. This allows the threats they make to be more effective and are therefore preferred platforms.
    - Uses an online platform account created for the purpose of sexual extortion or a hacked account that can provide authenticity.
    - Operates multiple accounts at any one time, due to the unverified nature of many online platform accounts.
    - Harvests profile pictures, images and content from other online platform accounts to provide authenticity.
    - May view linked accounts on other platforms to obtain sufficient information to carry out sexual extortion. The perpetrator may move the victim through multiple online platforms to harvest information.
    - Initiates communications of a sexual nature or with sexual overtones. They agree to send images of themselves, purporting to be someone the victim may be interested in, in various stages of undress.
3. The perpetrator may request communication moves to a different online platform to carry out the exchange of intimate images and threats to share them. This alternative online platform account may:
    - Not require as much resourcing and upkeep because the victim has been engaged and sufficient trust is built with the perpetrator.
    - Require little to no verification and know your customer requirements.
    - The perpetrator requests the victim provides images in various stages of undress.
    - The victim provides perpetrator images of themselves in various stages of undress.
    - The perpetrator advises the victim that they have harvested their information and then threatens to share their images if they don't give in to their demand, which is most commonly a financial demand. The perpetrator may also promise to delete the image, if they are paid.

- Even if the victim pays, the perpetrator will use the image as leverage for further demands. The threats will remain ongoing.

4. There is continued pressure on the victim to pay, which may include:
   - Opening group chats with contacts harvested from the victim's social media profile/s. They will then commence or threaten to commence sharing of sanitised or unsanitised images.
   - Using harvested information about social networks.
   - Using a countdown timer to increase time pressure.

Education is paramount in not only attempting to prevent sexual extortion but also building resilience and help-finding attitudes among the Australian community. eSafety publishes a range of information and guidance on its website relating to sexual extortion and generates advisories about topics of concern, such as sexual extortion.

*Digitally Altered Intimate Images*

The prominence of deepfake image-based abuse garners media attention from time to time. The definition of an intimate image in the OSA makes it clear that fake or digitally altered images, including AI deepfake material, are included.

eSafety has received a small number of reports involving digitally altered intimate images, comprising less than 1% of the overall number of reports, in annual reporting periods. One matter where eSafety is seeking civil penalties for the non-consensual sharing of digitally altered intimate images remains before the Federal Court of Australia, however, that matter was commenced in 2023.

We monitor developments in this area and proactively raise public awareness through outreach to the Australian community, educational materials on our website and the publication of position papers focused on generative artificial intelligence capabilities and deepfakes. We also proactively engage with industry to address emerging risks. As part of our broader response, eSafety has developed incident response guidance for schools, recognising the disruptive impact that deepfake abuse can have on students, families, and school communities.

**Addressing cybercrime through international engagement**

In 2025, bilateral engagements with key partners facilitated ongoing expert exchanges on thematic and operational priorities and developments, including:

- enhancing capabilities and knowledge of regulatory tools and approaches to industry supervision and enforcement
- deepening knowledge of systemic approaches to regulation, including developing and overseeing codes and rules, risk assessments and transparency reporting - particularly through collaboration with partners such as the OECD
- understanding issues such as protection of children and children's rights, social media minimum age (SMMA) and age assurance technologies
- understanding high risk and high priority harms (e.g. TVEC, child sexual exploitation and abuse (CSEA), and image-based abuse (IBA))
- understanding intersections with AI and other emerging technologies and identifying measures to respond at scale, and supporting regulatory activities by sharing insights into platform

features, functions, and incident trends, including through collaboration with INHOPE and other global networks

[redacted]

[redacted]

**Results of the Australian Cybercrime Survey trial**

eSafety partnered with the Australian Institute of Criminology and the Joint Policing Cybercrime Coordination Centre (JPC3) on a longitudinal trial of cybercrime intervention messages. The trial assessed the impact of providing preventative messaging via monthly educational emails regarding online abuse and harassment (eSafety) and profit-motivated cybercrime (JPC3).

The trial results revealed that this messaging model shows promise for raising awareness and understanding of eSafety and its offering, and encouraging preventative action, as summarised below:

*Engagement with messaging*

- Within the eSafety intervention group, there was an average email open rate of 51%, peaking at 64% at the start of the trial, indicating strong initial interest.
- 95% of participants found the messages easy to understand at least some of the time.
- 65% took action to improve online safety based on the messages some of the time; 24% did so most of the time.

*Awareness and understanding of eSafety*

- Participants exposed to the intervention were more likely to have heard of eSafety (82% vs. 61.4%) relative to the control group participants.
- They also had a better understanding of eSafety's role in preventing online harms (60.6% vs. 46.6%).

*Impact on online safety behaviours*

- The trial saw an overall decline in victimisation (online abuse, harassment, malware) between 2023–2024. There was no statistically significant difference in victimisation rates between intervention and control groups.

- Overall, unsafe behaviours increased between the two time points, but less so among those exposed to eSafety messaging.
- Furthermore, while the control group showed declines in both privacy and security behaviours, these remained stable amongst those exposed to eSafety messaging.
- Overall, the results are encouraging, and eSafety is interested in exploring further how this messaging modality might be harnessed to promote awareness of eSafety and engagement with safe online behaviours for different subgroups of the Australian population.

**Industry codes and standards**

The OSA provides for industry bodies to develop mandatory industry codes to regulate 'class 1' and 'class 2' material. Class 1 material includes CSEM and pro-terror content, while class 2 material relates primarily to adult content such as pornography. Codes are produced by industry associations per industry sector and then submitted to the eSafety Commissioner for registration. If a code meets the relevant legislative requirements, including that it provides appropriate community safeguards, then the eSafety Commissioner can decide to register it. If the code fails to satisfy that test, then the eSafety Commissioner can declare a standard for that industry sector.

There are currently six industry codes, and two industry standards in force which require the online industry to take steps to address the most harmful material (class 1A and 1B material), including CSEA and pro-terror material. The requirements apply to social media, messaging, gaming, dating, file sharing, generative AI services (including 'nudify' services and model distribution services), search and app stores, amongst others. Requirements include the detection of known CSAM, and disruption and deterrence of new CSEA, as well as requiring timely investigation of user reports, and sufficient trust and safety resourcing. eSafety is prepared to use the full range of our enforcement powers, which include formal warnings, infringement notices, and seeking a civil penalty from the Federal Court of up to $49.5 million (AUD).

A second phase of codes are focused on preventing children's access to age-inappropriate material (Class 2 material) such as online pornography and providing all users with effective information, tools and options to limit access and exposure to this material. This June, the eSafety Commissioner registered three of the nine draft Phase 2 codes submitted by industry, covering search engines, enterprise hosting and internet carriage services. Decisions on the remaining six draft codes – which cover social media, messaging, app distribution, equipment providers, online games, and websites – are under consideration.

I look forward to the opportunity to discuss these matters again with you and the Committee, should you wish to learn more about our ongoing efforts to address cyber-enabled crimes and harms online.

Yours sincerely,

Julie Inman Grant
eSafety Commissioner