eSafety Commissioner — Australian Government

15 December 2023

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 contribute to the Joint Committee on Law Enforcement's inquiry into the capability of law enforcement to respond to cybercrime, including cyber-enabled crime. We have concentrated on the following terms of reference:

c) Coordination efforts across law enforcement, non-government and private sector organisations to respond to the conduct of cybercrimes and risks of cybercrime
d) Emerging cybercrime threats and challenges affecting Australian entities and individuals, including the scale and scope of cybercrimes conducted in Australia or against Australians
f) Prevention and education approaches and strategies to reduce the prevalence of victimisation through cybercrime, and
g) other related matters.

eSafety is Australia's independent regulator for 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.

eSafety has a range of functions under the *Online Safety Act 2021* (Cth) ('OSA'). These include administering complaints schemes for:
- cyber-bullying material targeted at an Australian child
- non-consensual sharing of intimate images, also known as image-based abuse (IBA)
- cyber-abuse material targeted at an Australian adult, and
- illegal and restricted online content such as child sexual exploitation material (CSEM) and pro-terror material.

The OSA also provides tools to regulate services' systems and processes. These include enabling eSafety to require online service providers to report on the steps they are taking to comply with the Basic Online Safety Expectations, which outline the Australian government's expectations for certain types of online services to minimise material or activity that is unlawful or harmful. The Act also provides for representatives of sections of the online industry to develop new industry codes relating to the online activities of participants in those sections of the online industry. The industry codes are intended to regulate illegal and restricted content, including CSEM.

Other fundamental elements of our successful regulatory model include prevention through awareness and education and initiatives to promote proactive and systemic change through frameworks such as Safety by Design.

There can be a degree of overlap between the issues eSafety regulates through the OSA and certain forms of cyber-enabled crime. eSafety undertakes its work within the civil regulatory context, and so is functionally separate from law enforcement. Nevertheless, we regard our efforts as being

complementary to and supportive of Australian police, and the ways we work together are expressed in memorandums of understanding with policing organisations around the country (including the Australian Federal Police).

In most cases, the harms reported to eSafety can be resolved through the application of both formal and informal regulatory power. The objective of eSafety regulatory investigations is always to alleviate harm, whether through removal of harmful material or, when provided for by the OSA, restraining the ability of individuals to perpetrate harm. In addition to achieving harms reduction, eSafety will often also help to connect affected complainants to additional support, including services such as Kids Helpline.

When material or activity appears to meet a criminal threshold or where there is evidence that a person is under a real threat of physical harm, we work with law enforcement – including through referral of specific matters for criminal investigation – to achieve a degree of coordinated, cross-agency and multi-jurisdictional effort. This is especially true in cases where under 18s experience IBA, which we refer for assessment and triage to the Australian Centre to Counter Child Exploitation 'ACCCE').

Below, we provide examples of two forms of cyber-enabled crime which eSafety addresses through our civil regulatory framework under the OSA. The first is child sexual exploitation material, while the second is IBA (including sexual extortion). We employ a variety of tactics to address these harms, including through direct intervention to alleviate harm, driving education and prevention efforts to raise awareness and shift attitudes and behaviours, encouraging proactive and systemic change through Safety by Design, and administering industry regulation powers to lift safety standards across the online tech sector.

Illegal and Restricted Content reports
eSafety investigates complaints from the public about CSEM, pro-terror content and other forms of illegal or harmful online content online. In FY 2022-23, eSafety received complaints about 33,000 URLs, more than twice the number received in FY 2021-22. About 87% of all reports concerned CSEM.

Tackling CSEM requires a cross-jurisdictional effort, as very little is hosted in or provided from Australia. To this end, eSafety is a member of the global network formed by the International Association of Internet Hotlines (INHOPE), comprised of more than 50 organisations around the world that achieve rapid removal of CSEM, in tandem with industry and law enforcement. In rare cases where an Australian connection is observed, eSafety will coordinate with police from the relevant jurisdiction using protocols established in MOUs.

The prevalence and accessibility of CSEM online continues to be a pervasive concern requiring strategic disruption responses extending beyond the criminal justice system. Addressing the many elements that enable the online sexual exploitation of children demands a whole-of-government, whole-of-community approach that reaches across borders.

The 2023 Global Threat Assessment by the WeProtect Global Alliance (which includes the eSafety Commissioner on its board) highlights the scale and escalation of online child sexual exploitation and abuse worldwide as it reports the volume of child sexual abuse material reports to the US National Center for Missing & Exploited Children (NCMEC) increased by 87% since 2019. The WeProtect Global Alliance also reports that new forms of online abuse, like financial sexual extortion and AI-generated imagery pose new challenges, underscoring the urgent need for Safety by Design.

Image-based abuse reports and sexual extortion
The Image-based Abuse Scheme consists of eSafety's regulatory powers to investigate and provide direct assistance to individuals whose intimate images or videos have been shared online (or threatened to be shared) without their consent. Most complaints about image-based abuse to eSafety concern offenders coercing Australians into producing intimate images of themselves and then extorting them for

payment (sexual extortion).

Since the COVID-19 pandemic, there has been a sharp rise in the number of complaints to eSafety about sexual extortion. In FY 2020-21, eSafety received about 2,700 complaints about image-based abuse. However, this number climbed by more than 55% the following year to 4,196 complaints, driven in part by sexual extortion. In FY 2022-23, eSafety received more than 9,000 IBA complaints, of which about 80% consisted of sexual extortion reports.

We understand through our engagement with law enforcement and from media reports that authorities globally are seeing a significant increase in offshore criminal syndicates targeting children and young adults in this way. The cost to these young people is significant. Many have paid large sums and countless have suffered deep distress, with some reports here and overseas of young people taking their lives as a result.

Offenders have a well-established method of attack. Typically, the target is a male aged between 18 and 24 years, and offenders often make initial contact with them on a platform such as Instagram. Offenders generally employ fake accounts featuring profile images of attractive young women. A conversation with the target ensues, following which the offender encourages the target to add them to a secondary service (for example, Snapchat). Once on that secondary service, the target is encouraged to produce intimate material and share it with the offending account, at which point the traps shuts and a demand for payment is made, accompanied by a threat to share the image or video with the target's social network if they refuse.

The sheer scale of sexual extortion in Australia means that it cannot be solved by investigating individual reports. We believe that prevention through education and awareness is likely to be the most effective solution to this problem, providing targets with the information and support they need to take back control. To this end, we are taking a collaborative, multi-agency approach to our prevention efforts. For example, we are developing a communication and education strategy with our law enforcement partners to simply express two messages: that targets not pay the offender (lest they become the 'bank' and so exacerbating the harm) and that they tell someone (parents, guardians, carers, friends, or support services).

Another limb of our prevention response involves systemic disruption through engagement with industry (particularly with Instagram and Snapchat, the most highly enabling platforms) and intelligence sharing with Commonwealth agencies, including the AFP (via the ACCCE) and the Australian Transaction Reports and Analysis Centre (AUSTRAC).

We have established a close working relationship and agreed processes with our partners at the AFP-led ACCCE to respond to reports to eSafety from Australian children and young people under 18 years experiencing sexual extortion. eSafety plays a part in the Australian Government's overall response to these harms, and routinely attends roundtable and information-sharing meetings convened by the Attorney General's Department with the ACCCE and key mainstream online service providers.

Addressing cybercrime through international engagement
Protection of children online is now a main feature of many United Nations (UN) and multilateral forums. eSafety works with the Department of Foreign Affairs and Trade (DFAT) to advance Australia's core priorities through the Commission on Crime Prevention and Criminal Justice, including to counter cybercrime, such as the online abuse and exploitation of children. eSafety is also contributing to whole of government consultations to support DFAT's negotiations on the proposed UN Cybercrimes Convention. We help lead global efforts to protect children online through the eSafety Commissioner's position on the board of the WeProtect Global Alliance and our involvement in the Alliance's Global Task Force. In late 2022, eSafety launched the Global Online Safety Regulators Network to promote cooperation and collaboration among online safety regulators, with members and observers to the Network now spanning

five continents. Recently, eSafety handed over our position as inaugural Chair to Ofcom.

We also play a key role in the INHOPE network and are close partners of the UK's Internet Watch Foundation and the Canadian Centre for Child Protection. eSafety also delivers capacity building projects in the Indo-Pacific region under DFAT's Cyber and Critical Cooperation Program, with a focus on child online safety and technology-facilitated gender-based violence.

Prevention and education efforts

eSafety has a legislated role under the OSA to improve and promote online safety for Australians, which includes supporting and encouraging online safety education in Australia.

We regard the eSafety website (www.esafety.gov.au) as an essential tool in our mission to help Australians affected by online harms – including those outlined above. Using the website, Australians can easily find extensive advice about dealing with unwanted contact and grooming, the steps needed to safely respond to sexual extortion threats and attempts, and ways to manage hard-to-have conversations with children about online safety. We also design and deliver training to key audiences with direct influence over children and young people, including educators, parents, and carers. Other audiences include law enforcement, family and domestic violence frontline workers and senior Australians through the award-winning BeConnected program.

eSafety also oversees a range of specifically funded programs. These include the Families Capacity Building project, which aims to extend eSafety's ability to convey online safety information to families that are harder to reach, and the Children and Tech Facilitated Abuse project for children impacted by technology abuse occurring in family and domestic violence settings. Other programs include the development of a support service related to tech-based family, domestic and sexual violence and programs that enable us to reach at-risk audiences through upskilling organisations and those on the frontline, such as sporting administrators and athletes.

Australian Cybercrime Survey trial

eSafety has 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 will assess the impact of providing preventative messaging via monthly educational emails regarding online abuse and harassment (eSafety) and profit-motivated cybercrime (JPC3). The trial offers eSafety and our research partners an opportunity to test, using a rigorous research design:

- the impact of prevention messages on victimisation, responses to victimisation, and engagement with risky and protective behaviours,
- the level of engagement of different subgroups of the population with varying message content, and
- the overall efficacy of this message modality.

Safety by Design

Recognising the importance of embedding safety measures from the start and throughout the design, development and deployment processes of digital products and services, eSafety introduced the Safety by Design initiative in 2018. This involved conducting in-depth research and engaging in consultations with industry stakeholders. Safety by Design serves as a framework for technology companies to proactively address safety concerns and foster a safer online environment for their users.

The initiative is underpinned by three principles covering service provider responsibility, user empowerment and autonomy, and transparency and accountability. The principles have been translated into a set of comprehensive risk assessment 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.

eSafety stays ahead of issues related to emerging technologies such as generative artificial intelligence and immersive environments, through ongoing consultation with a wide variety of industry and academic stakeholders, and via horizon scanning. This proactive approach identifies potential areas of concern arising from rapid developments in emerging technologies, promotes best practices for safe product design and development across industries, and positions Safety by Design as an initiative which can enable technology companies to anticipate, detect and eliminate online risks to make our digital environments safer and more inclusive, especially for those most at risk.

Regulating industry through codes, standards and transparency

In addition to the direct powers provided under the OSA to address individual harms such as image-based abuse and the distribution of CSEM online, the eSafety Commissioner can also regulate certain aspects of the online industry via mandatory industry codes and the Basic Online Safety Expectations.

Basic Online Safety Expectations

The OSA provides eSafety with powers to require online services providers to report on the reasonable steps they are taking to comply with the Basic Online Safety Expectations (BOSE). The obligation to respond to a reporting requirement is enforceable and backed by civil penalties and other mechanisms. eSafety can also publish statements about the extent to which services are meeting the Expectations. The BOSE are determined by the Minister for Communications and set out the Australian Government's expectations of certain kinds of online services. Examples of these expectations include that providers are:

- taking reasonable steps to proactively minimise material or activity that is unlawful or harmful, and ensuring users can use a service in a safe manner
- protecting children from content that is not age appropriate like pornography
- taking reasonable steps to prevent harmful use of anonymous and encrypted services
- putting in place user-reporting mechanisms, and clearly outlining their terms of service and enforcing penalties for people who breach these terms
- cooperating with other service providers
- responding to requests for information from the eSafety Commissioner.

The requirements are designed to improve providers' safety standards, and to improve transparency and accountability. Using the power, eSafety aims to lift the hood on what services are and are not doing to prevent and detect abuse. eSafety has issued twelve notices in the last year to some of the largest technology companies in the world (including Google, Meta, Twitter/X Corp. and Microsoft) focussed on child sexual exploitation and abuse. Through the data obtained from these notices, eSafety published two transparency reports that showed, for example, that many companies were not taking even relatively simple steps to protect children and failing to use widely available PhotoDNA technology to detect and remove child abuse material.

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. Under the OSA, codes must be registered across eight industry sectors, including providers of social media services, internet service providers, and enterprise hosting providers.

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 represents appropriate community safeguards, then the 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.

The industry bodies tasked with developing the industry codes adopted a two-phase approach, as suggested in eSafety's 2021 position paper. This approach prioritised the most harmful material for regulation through the codes, with the first phase focusing on the most harmful types of class 1 content, including child sexual exploitation material and terrorist material.

Five industry codes addressing this material were registered by the eSafety Commissioner on 16 June 2023, and come into effect on 16 December 2023. They cover the following industry sectors:
- social media services
- internet carriage services (also known as internet service providers)
- equipment providers
- app distribution services
- hosting services.

A code covering how internet search engine services will deal with such material, including the risk presented by integrated generative AI capabilities, was registered on 12 September 2023 and will come into effect on 12 March 2024.

The codes set out a range of mandatory compliance measures that are legally enforceable. For example, the largest social media services (known under the social media services code as 'Tier 1' social media services) are required to deploy systems, processes and/or technologies to detect instances of 'known' (pre-categorised) child sexual abuse material. Where a company (such as the provider of a social media service) is not complying with an industry code, the eSafety Commissioner can direct the provider to comply. A failure to comply with such a direction is a civil penalty provision.

Two of the draft codes submitted by industry groups for registration, relating to Designated Internet Services ('DIS', including websites and end-user managed hosting services) and Relevant Electronic Services ('RES', including messaging and email services, and online gaming services) were not registered by the eSafety Commissioner, as they failed to meet appropriate community safeguards. For example, the DIS code did not commit to the detection by providers of end-user managed hosting services of known child sexual abuse material and the RES code did not require email services and some partially encrypted messaging services to do the same. The eSafety Commissioner has now drafted industry standards for these two sectors to rectify the draft codes' shortcoming. These draft standards are currently open for public consultation.

I look forward to the opportunity to discuss these and other matters with you and the Committee, should you wish to learn more about the eSafety Commissioner's role and functions in combatting cyber-enabled crimes and harms online.

Yours sincerely,

eSafety Commissioner