Harnessing data and digital technology interim report: eSafety submission

The eSafety Commissioner (eSafety) welcomes the opportunity to provide a submission to the Productivity Commission's *Harnessing data and digital technology* interim report.

eSafety is Australia's independent regulator, educator and coordinator for online safety. Our purpose is to help safeguard Australians from online harms and to promote safer, more positive online experiences.

We undertake a range of work to understand, anticipate and respond to online harms and promote online safety. This includes in relation to digital technologies like artificial intelligence (AI). We also share our expertise and experience as Australia's online safety regulator to ensure our in-depth understanding of online safety and digital technology can inform related measures and initiatives across Australia.

In addition to this submission, we are pleased to have co-authored a joint submission to the *Harnessing data and digital technology interim report* as part of the Digital Platform Regulators Forum (DP-REG). It outlines how DP-REG members are working both individually and together to help Australians safely realise the productivity gains made available by AI, as well as the role DP-REG could play to uplift the tools and capabilities of other regulators to regulate AI effectively.

Safety as an enabling factor

This submission highlights how strong safeguards, including in relation to online safety, enable economic growth and broader public good, as they build trust and create an environment that enables productivity and promotes innovation.

eSafety supports the need to ensure the benefits of data and digital technologies, including in relation to AI, can be realised, while managing and mitigating harms and downside risks. We also support an approach to AI that balances and prioritises both opportunity and responsibility, building upon existing regulatory frameworks and consistent with eSafety's Safety by Design initiative.

In summary, we see:

- Safety as an innovation driver: embedding safety boosts trust and system performance, as well as creating a culture of continuous improvement and innovation.
- Safety as an economic enabler: online harms reduce productivity and undermine consumer trust, whereas safety improves engagement and adoption.

- Safety as an enabler for new markets: safety can enable new markets for competition and provide opportunities for organisations to differentiate themselves by innovating on safety features.
- Safety as a means of digital inclusion: safety ensures equitable and inclusive experiences for the whole community, including at-risk groups.
- Safety to build trust: it can take years to build trust and a moment to lose it. Embedding safety throughout the AI lifecycle builds confidence and sends a signal to users that encourages and supports their engagement.
- Safety as a basis for cross-government alignment: online harms intersect with a range of remits, including privacy, security and AI reforms.
- Safety through AI: AI technologies like abuse detection tools can help address harm and promote safety, as well as privacy and security.

Regulation of AI

While the benefits of AI are vast and growing, they won't be realised unless appropriate action is taken to address risks and harms. This includes establishing the right regulatory settings. This requires effective, balanced and proportionate regulation, supported by policy, procedure and standards.

Overall, eSafety supports a coordinated, whole of government approach to AI regulation that uses existing laws and regulatory structures to minimise harms. This will help ensure that the subject matter specificity and expertise that has emerged in the context of existing regulatory remits is maintained, while a whole of government approach also works to minimise fragmentation and inconsistency.

We believe the starting point in considering how to regulate AI should be to review existing frameworks. If gaps are identified, the next step should be to strengthen existing frameworks, rather than pursue a standalone legislative approach specific to AI. As we explore later in this submission, the *Online Safety Act 2021* (Online Safety Act) has recently been independently reviewed. The Final Report was provided to the Australian Government in October 2024. It made 67 recommendations to strengthen Australia's online safety legislative framework and ensure it is effective, up to date and future proofed.

In relation to AI-specific regulation, we note tech-specific regulation has generally proved ineffective and cumbersome. This is why newer approaches promote technology-agnostic or technology-neutral considerations.

eSafety considers AI-specific regulation presents practical implementation challenges, in addition to consumer and industry confusion and regulatory burden. We also note the risk that if AI-specific regulation is developed, it may lead to other tech-specific regulation that fragments regulatory approaches and fails to adequately cover emerging convergence with other technologies. The rapidly evolving nature of digital technology means regulatory responses to AI are best addressed under existing frameworks, which are proportionate, risk-based, outcomes-based and technology-neutral.

In addition to legislation being flexible enough for regulators to respond to emerging technology and harms, it is imperative that regulators have the right tools, capabilities and expertise to respond.

eSafety supports consideration of society-wide approaches to digital technologies and the digital economy. We also highlight the need to consider the impacts at an individual level. Without deliberate action, digital technologies risk reinforcing existing social inequalities – a concern that is further intensified by the potential for emerging technologies, such as artificial intelligence, to amplify existing risks and harms. To address this, it is essential to embed digital literacy and capability development, particularly in relation to AI, across the broader Australian community as part of a comprehensive, society-wide approach to the digital economy.

This will ensure the harms of digital technology do not disproportionately impact underrepresented and marginalised communities, while ensuring digital technology benefits all Australians.

eSafety's approach and existing framework

In line with our view that reviewing and, where necessary, strengthening existing frameworks is preferrable, we outline eSafety's existing approach and regulatory measures to address AI. We then outline existing reform that is underway in relation to the Online Safety Act, noting this is the best basis for pursuing any reform relating to AI in the context of online safety.

Our regulatory approach is underpinned by three pillars of prevention, protection, and proactive and systemic change.

These pillars reflect our broad and holistic remit. The way the pillars work together reflects how eSafety's various functions work together to create a multidimensional regulatory toolkit.

We take a risk and harms-based approach to our work. As noted above, this is also strengths-based and adopts an intersectional lens.

We have a broad regulatory remit that gives us a range of regulatory tools under the Online Safety Act relating to AI material.

Complaints Schemes

A key underpinning of eSafety's regulatory approach is that we are a safety net for individuals.

Under the Online Safety Act, eSafety administers complaints schemes to address:

- Cyberbullying material targeted at Australian children.
- Non-consensual sharing of intimate images (image-based abuse).
- Cyber-abuse material targeted at Australian adults.
- Illegal and restricted online content, such as child sexual abuse material and pro-terror material.

eSafety can investigate and facilitate removal of both real and synthetic material. For example, this means we can seek removal of child sexual abuse material or image-based abuse material, even if it has been digitally altered or generated by AI.

Codes and standards

Our systemic regulation, such as industry codes and standards, creates mandatory obligations for eight key sections of the online industry.

Industry codes and standards under the Online Safety Act address specific online safety issues in eight sections of the online industry across the digital stack. While AI-generated material is treated under the legislation in the same way as 'real' class 1 and class 2 material, the risks associated with AI generated material have necessitated specific requirements in relation to AI-related features in some codes and standards.

We enforce eight Industry Codes and Standards that deal with illegal and harmful content, two of which contain measures specific to generative AI: the industry code for search engine services and the industry standard for designated internet services, which applies to a broad range of apps and websites, such as ChatGPT and AI enabled applications like 'Nudify' apps

and AI companion apps. All the codes and standards apply to material whether it is artificially generated or not.

The Commissioner has recently registered additional industry Codes that will come into effect over the course of late 2025 and 2026. Measures in these Codes include specific measures to address how children may use generative AI services to generate sexually explicit imagery, and to create protections for children using AI companion chatbots. They will apply to the same range of services as the existing Industry Codes and Standards.

Basic Online Safety Expectations

We can also compel transparency from certain services under the Basic Online Safety Expectations (BOSE). The BOSE outline the Australian Government's expectations that social media, messaging and gaming service providers and other apps and websites will take reasonable steps to keep Australians safe.

In May 2024, the BOSE were expanded to include an explicit expectation related to generative AI. There are also measures that are likely to require the proactive use of AI to prevent and detect harmful and unlawful material, such as new child sexual abuse material. We are also empowered to compel transparency from industry by requesting information about how services are meeting the BOSE.

eSafety has issued transparency notices to industry seeking to understand how AI is used to improve safety on online services, including how services are using AI to detect and remove child sexual abuse material and grooming. We've also questioned how services are addressing safety risks of AI, such as the potential amplification of harmful content via recommender systems. These findings are published on eSafety's website.

In March 2025, eSafety published a summary of the findings of the notices issued in March 2024 covering, among other things, generative AI in relation to terrorism and extremism, and child sexual abuse. This report highlighted that services such as Google's Gemini were already receiving user reports of suspected AI generated terrorist and violent extremist material.

In July 2024, the first 'periodic' notices were issued to eight providers. These require four six-monthly reports to eSafety for a period of two years, relating to child sexual exploitation and abuse, including sexual extortion of children and grooming, as well as sexual extortion of adults. eSafety asked specific questions of providers that offered generative AI features on their services. In August 2025, eSafety published its findings from the providers first responses to the transparency notice.

Safety by Design

Through our Safety by Design initiative, we work with tech companies to shift their design ethos from 'moving fast and breaking things' to moving thoughtfully and anticipating, detecting and eliminating online threats. This is done by assessing risks and potential harms, thereby embedding protections at the front end, and throughout the design, development and deployment processes. This includes the assessment of harms and threats posed by AI, the proactive mitigation of potential misuse and constant monitoring for unanticipated harms following deployment. By creating safer user experiences from the start, a Safety by Design approach centres the user and helps tech companies to scale responsibly.

Research

A central objective of our research program is to generate robust, person-centred evidence on the prevalence, nature and impact of online harms in the Australian community. This includes dedicated research into online harms, trends and insights, informed by national surveys, analysis of data from eSafety's reporting schemes and collaboration with external subject matter experts. This research aims to strengthen the evidence base to inform prevention, policy, regulatory and service responses, and to ensure that stakeholders have access to timely, actionable insights. Given the extent to which eSafety's remit captures AI, our research often explores the use of AI and AI-generated harms.

For example, eSafety's 'Keeping Kids Safe Online' survey found that over 2 in 5 children aged 10-17 (42%), living in Australia, had ever used generative AI (eSafety, Forthcoming)[1]. We are also developing a survey to determine the prevalence and nature of AI assistant and companion app usage among children aged 10-17, living in Australia.

Policy and tech analysis and environmental scanning

As part of our Tech Trends program, we conduct horizon scanning and work with subject matter experts to identify the online safety risks and benefits of emerging technologies, as well as the regulatory challenges and benefits they may present. This enables eSafety to be anticipatory, both with respect to harnessing the benefits of emerging technology, but also in mitigating potential risks. Our published outputs relating to AI-generated online harms under this program include:

[1] eSafety Commissioner. (2025). *Connected, curious, cautious: Children's engagement in the digital world*. Australian Government.

- Our position statement on generative AI providing an overview of the generative AI lifecycle, examples of its use and misuse, consideration of online safety risks and opportunities, and regulatory challenges and approaches.
- Our position statement on recommender systems and algorithms exploring the benefits and risks of recommender systems and their underlying algorithms, as well as impacts on an individual and societal level.
- A blog post in June 2025 discussing the convergence of generative AI and child sexual exploitation and abuse.

We undertake broad ranging policy analysis and advice work. This includes our submissions to various inquiries dealing with AI, such as our:

- Submission to the *Introducing mandatory guardrails for AI in high-risk settings: proposals paper* (Mandatory Guardrails Paper) (October 2024).
- Submission to the Senate Standing Committee on Legal and Constitutional Affairs Legislation Committee on the *Criminal Code Amendment (Deepfake Sexual Material) Bill 2024* (July 2024).
- Submission to the inquiry into the use of generative AI in the Australian education system (July 2023).

Education and awareness raising

Our education, awareness raising and engagement efforts relating to online safety include education relating to the safety impacts of AI. We are already delivering webinars for educators and parents and carers about aspects of AI, including on AI assisted image-based abuse and understanding AI companions.

AI has also been discussed in some of the key advisory and consultation mechanisms we oversee, including the eSafety Youth Council, the Trusted eSafety Providers and the National Online Safety Education Council.

We publish a range of advisories related to AI. For example, in February 2025 we published an advisory on AI chatbots and companions – risks to children and young people, and in June 2025 we published an advisory on Deepfake damage in schools.

Underpinning all our education and awareness raising work is the importance of digital literacy. Digital literacy is the foundation for people to be able to able to interact online safely, thoughtfully, competently and critically.

International engagement

We recognise that online safety is a global challenge that requires a collaborative response to achieve a greater degree of regulatory coherence. Through our international engagement and collaboration, we work with online safety regulators around the world. This includes through the Global Online Safety Regulators Network (GOSRN) and various other international forums and networks. We engage in regular cross-jurisdictional and multi-disciplinary dialogue with our international counterparts and colleagues to better understand regulatory best practice, improve regulatory coherence and protect Australians in the global online environment.

Existing and ongoing online safety reform

As noted above, the Online Safety Act was recently independently reviewed, with the Final Report provided to the Australian Government in October 2024.

In November 2024, the Australian Government committed to legislating a duty of care. While details of the model are still being developed, it will place the onus on digital platforms to proactively ensure online safety, such as by taking reasonable steps to prevent foreseeable harms on their services. We also expect the duty of care will result in an obligation on industry to implement stronger protections against online harms at a systemic level. The duty of care will be able to draw upon the work eSafety has done to date, including our systemic work and Safety by Design initiative, which has already been incorporated into many of the industry codes and standards registered by eSafety.

eSafety will continue to support the Government in developing the duty of care, as well as progressing any other reform identified from the recent comprehensive review.

Conclusion

eSafety is pleased to have the opportunity to contribute to the Productivity Commission's important inquiry. In summary, strong safeguards, in existing legislation and through a Safety by Design approach, are the foundation for trust, innovation and productivity.

eSafety supports the need to ensure the benefits of data and digital technologies, including AI, can be realised, while managing and mitigating harms and downside risks. We support an approach to AI that balances and prioritises both opportunity and responsibility and promotes a whole of government approach to AI that ensures particular policy areas, such as online safety, retain their cohesiveness and subject matter specificity. We are happy to provide any further information that would assist.