DP-REG Joint Submission to Department of Industry, Science and Resources – 'Safe and Responsible AI in Australia' discussion paper

1.1. The Digital Platform Regulators Forum (DP-REG) welcomes the opportunity to contribute to the Department of Industry, Science and Resources (DISR) consultation on the 'Safe and responsible AI in Australia' Discussion Paper (the Discussion Paper).

1.2. DP-REG is an information-sharing and collaboration initiative between Australian independent regulators with a shared goal of ensuring Australia's digital economy is a safe, trusted, fair, innovative and competitive space.

1.3. Emerging technologies such as artificial intelligence (AI) present new opportunities but also new challenges. An effective approach to the regulation of AI requires collaboration and coordination between regulators given the need for complementary expertise to address these challenges.

1.4. The purpose of this submission is to outline how DP-REG members are working together to understand the potential impacts posed by AI in Australia, and how our respective regulatory frameworks currently apply to AI technology.

2. About DP-REG

2.1. In March 2022, the Australian Communications and Media Authority (ACMA), the Australian Competition and Consumer Commission (ACCC), the Office of the Australian Information Commissioner (OAIC), and the eSafety Commissioner (eSafety) formalised existing collaborative arrangements to form DP-REG.

2.2. Through DP-REG, members share information about, and collaborate on, cross-cutting issues and activities involving the regulation of digital platforms. This includes consideration of how competition, consumer protection, privacy, online safety and data issues intersect. The structure, purpose and goals of DP-REG are outlined in our Terms of Reference.

2.3. This forum is similar to bodies set up in other jurisdictions such as the Digital Regulation Cooperation Forum (DRCF) in the United Kingdom[1] and the Digital Regulation Cooperation Platform in the Netherlands.[2]

2.4. DP-REG's strategic priorities for 2023-24, as outlined in our July 2023 communique, include assessing the impact of algorithms, seeking to improve transparency of digital platforms' activities and how they are protecting users from potential harm; increased collaboration and capacity building between the four members; and a new focus on understanding and assessing the benefits, risks and harms of generative AI.

---

[1] DRCF, Digital Regulation Cooperation Forum, accessed 5 July 2023.

[2] Authority for Consumers & Markets, Digital Regulation Cooperation Platform, accessed 5 July 2023.

2.5.   Working groups progress the key priorities, projects and activities of DP-REG. Currently, DP-REG has three standing working groups:

- **Digital Technology Working Group** to jointly explore relevant digital platform technologies (including algorithms) and their regulatory implications

- **Codes & Regulation Working Group** to undertake activities that promote a consistent and coordinated approach to regulatory frameworks and common regulatory issues, and to build regulatory capability across DP-REG members

- **Data & Research Working Group** to undertake activities that support the collection and sharing of relevant data, research and information across DP-REG members.

2.6.   DP-REG's Digital Technology Working Group conducts joint work, such as a project evaluating the risk posed by algorithms and a 'technology examination' working paper on Large Language Models (LLMs). We expect this work will inform broader government processes, including this consultation and work underway by the Department of Infrastructure, Transport, Regional Development, Communications and the Arts, and the Department of Home Affairs in response to recommendations 13 and 14 of the House of Representatives Select Committee on Social Media and Online Safety. Each DP-REG member is also separately considering more specific harms stemming from AI relevant to their respective mandates.

3.   Developing an approach to AI governance

3.1.   The increasing adoption of AI – in particular, generative AI – could have broad-ranging benefits and risks for Australia's economy and society. As discussed below, immediate impacts of this technology include risks to consumer protection, competition, media and the information environment, privacy and online safety.

3.2.   The Discussion Paper notes that any regulatory and governance response to address the risks associated with AI should start by considering the extent to which Australia's existing regulatory frameworks provide appropriate safeguards.

3.3.   We support this approach and, where gaps are identified, the Government should consider how existing frameworks may be strengthened and enhanced (including through existing regulatory reform proposals) before consideration is given to creating a separate regime specific to this technology.

3.4.   The effective coordination between DP-REG members, as well as other arms of Government, will therefore be crucial to the development of effective regulatory approaches to AI. Through DP-REG, its members engage in ongoing collaboration, information sharing and coordination on matters relating to digital platforms regulation, including engagement with Government counterparts, academic experts, and industry stakeholders. By continuing this work, the forum will be able to make a valuable contribution to whole-of-Government discussions about Australia's response to AI.

3.5.   The section below sets out how the existing laws and regulatory powers of DP-REG members, and other ongoing law reform processes, may operate to address current or emerging risks associated with AI, and where potential gaps may lie.

4. Application of the existing regulatory frameworks to AI

4.1. The use of AI by digital platforms has impacts on users, businesses, and government. Many of these impacts may exacerbate existing, already widespread risks that digital platform regulators are already working to address.

4.2. We also note that the emergence and popular use of AI technology may pose issues that more broadly affect the ability of Australian regulators to exercise their responsibilities. For example, more widespread use of AI by regulated entities may highlight challenges regulators already experience in using their investigative powers to access algorithms, code and other technical material, which may be stored in other jurisdictions. Further, the ability of generative AI to produce large bodies of unique text may be misused to frustrate public submission processes run by regulators and other government agencies, putting a strain on staff time and resources, and making it difficult to accept and consider public submissions made in good faith.[3]

Consumer protection

4.3. The ACCC's consumer protection role includes enforcement of the Australian Consumer Law (ACL) to ensure that consumers and small businesses are protected from misleading and deceptive conduct, unconscionable conduct, unfair terms and conditions and unsafe products, and to promote fair trading. The ACCC also operates the National Anti-Scam Centre (NASC) and Scamwatch website which helps Australians learn how to recognise, report, and protect themselves from scams.

4.4. The ACCC has been considering the consumer impacts of digital platforms for a number of years, including in the 2019 Digital Platforms Inquiry (2019 DPI), and the interim reports of the Digital Platform Services Inquiry (DPSI). There have been 6 published reports to date, with two further reports underway.

4.5. While new products and services powered by generative AI have significant potential to benefit consumers and support productivity, this technology may also present new risks, or exacerbate existing risks to consumers online.[4]

Fake reviews, scams and harmful applications

4.6. As noted in the ACCC's DPSI reports and 2019 DPI Final Report, poor experiences online – due to scams, fake reviews and harmful applications – can harm individual consumers and broadly erode consumer trust in the digital economy.

4.7. While generative AI may help identify online scams quickly and assist with scam disruption,[5] it also has the potential to increase the volume, sophistication, and impact of scam activity and allow better targeting of scams across communication channels – including digital platforms, phone and SMS.[6] Generative AI may also

---

[3] GW Regulatory Studies Center, Will ChatGPT Break Notice and Comment for Regulations?, 13 January 2023.

[4] For example, the US Fair Trade Commission is investigating whether OpenAI engaged in unfair or deceptive practices, resulting in reputational harm to consumers, through data collection and the publication of false – see Washington Post, The FTC is investigating whether ChatGPT harms consumers, 13 July 2023.

[5] ACS, AI can detect scam calls in real time, 10 May 2022.

[6] For example, scammers can use generative AI to feed genuine messages into models to create text that convincingly impersonates trusted organisations (see Norton, Special Issue Norton Cyber Safety Pulse Report – The Cyber Risks of ChatGPT, 2 March 2023), or to generate emails targeted at specific groups or individuals using relevant keywords (see New York Times, Lina Khan: We Must Regulate A.I. Here's How, 3 May 2023). Scammers can also use bulk aggregation and analysis of scam data to help scammers write more convincing scams, and better target their scams.

be used to increase the volume and sophistication of fake reviews online, which can frustrate consumer choice and distort competition.[7]

4.8. There are currently no specific laws to identify and block scams perpetrated over digital platforms or 'over-the-top' online services. The ACMA's actions are currently limited to services regulated under the *Telecommunications Act 1997.*

4.9. The ACMA has powers to combat scams delivered by phone and SMS. For example, the ACMA has registered and enforces the Reducing Scam Calls and Scam SMS industry code,[8] which requires telecommunications providers to identify, trace and block scam calls and text messages. These rules can assist to identify and prevent phone and SMS scams that utilise generative AI.

4.10. The Government is currently considering the ACCC's September 2022 interim report of the DPSI, which recommends addressing the prevalence of scams, fake reviews and harmful applications through new mandatory processes including notice-and-action processes, reporting processes, verification of certain business users, and dispute resolution processes. It also recommended a new independent ombuds scheme to resolve disputes between digital platforms and consumers, including small businesses.[9]

4.11. The implementation of these recommendations would complement the recently-established NASC within the ACCC, and any potential code/s for banks, telecommunications providers and digital platforms involved in the scam supply chain.

4.12. The NASC will work together with government and other regulators, industry, law enforcement bodies and community organisations to improve information sharing and disrupt scam activity, including in relation to generative AI scams. The NASC builds on the work of the ACCC's Scamwatch service and will raise consumer awareness about harmful scams, making it easier to report scams and support the work of law enforcement and government agencies such as ACMA and the Australian Security Investments Commission (ASIC). The ACMA is a member of the Regulator Steering Group set up to support the planning of the NASC.

Product safety, misleading and deceptive conduct and unfair trading practices

4.13. While AI has the potential to enhance product safety outcomes for consumers – such as by detecting potential safety issues – it also raises new safety risks.

4.14. The ACCC is actively involved in discussions in international fora on how to promote safe AI design and the potential use of AI by consumer regulators. Discussions have also included challenges AI poses to allocating liability – for example, when products such as smart home systems are made unsafe by software updates.

4.15. LLMs can provide false but authoritative-sounding statements that could mislead users, including when consumers are making purchasing decisions. Additionally, the increasing popularity and 'hype' surrounding LLMs may incentivise spurious

---

[7] See, for example, ACCC, Digital platform services inquiry – September 2022 interim report – Regulatory reform, September 2022, p 8.

[8] ACMA, Action on scams, spam and telemarketing: January to March 2023, accessed 26 June 2023.

[9] ACCC, Digital platform services inquiry – September 2022 interim report – Regulatory reform, September 2022, chapter 4.

and misleading claims about the capabilities (or existence of) of AI technology in a wide range of products.[10]

4.16. The ACL applies to all products or services (except financial products and services), and contains prohibitions on misleading or deceptive conduct, and false or misleading representations. Similar prohibitions apply to financial products and services and are enforced by ASIC under the *ASIC Act 2001* (Cth).

4.17. However, as noted in the ACCC's September 2022 DPSI interim report, existing laws do not always adequately address online harms. As such, the report recommends a range of reforms to address these harms, including the introduction of an economy-wide prohibition on unfair trading practices (which would also address similar offline harms). We understand consideration of this possible reform is currently being progressed by the Government.

4.18. With the growing use of AI in consumer products, the ACCC also notes the application of the ACL to digital products (including AI products, and products using AI in their design and/or supply) could be set out more clearly.

4.19. The ACCC also continues to recommend including an explicit legal obligation in the ACL requiring businesses to supply safe consumer products and services (irrespective of whether AI is involved in their design or supply).

Competition

4.20. The other key mandate of the ACCC is to promote competition by enforcing the *Competition and Consumer Act 2010* (Cth), regulating national infrastructure (such as telecommunications infrastructure), implementing the Consumer Data Right, and undertaking market studies as directed by the Treasurer, including in relation to digital platforms services.

4.21. Effective competition in markets encourages firms to innovate and improve the value of their offerings to consumers, leading to more choice, lower prices, and higher quality products and services. The ACCC has extensively considered the competition issues in markets for digital platform services in the 2019 DPI, the Digital Advertising Services Inquiry and the interim reports of the DPSI.

4.22. Technological changes, such as the integration of generative AI into digital platform services, can lead to innovative new products and services. For example, the incorporation of ChatGPT into Microsoft's Bing search service enables AI-assisted answers in response to a user's search queries.[11] A key challenge is to ensure the field of generative AI remains innovation intensive.

4.23. However, the development and supply of generative AI systems, and their integration into digital platform services, can also raise many of the same barriers to entry and expansion that make some digital platforms tend towards concentration and could in fact magnify the potential for these effects to occur.

4.24. In particular, while open-source training data for general LLMs is available through digital libraries, firms with control over valuable or unique data may have an incentive to create or entrench a 'data advantage' by actively restricting access to that data. For example, Reddit recently announced new charges for Application Programming Interface (API) access to prevent firms training LLM models on its

---

[10] FTC, Keep your AI claims in check, 27 February 2023.

[11] Microsoft, Confirmed: the new Bing runs on Open AI's GPT-4, 14 March 2023.

library of public posts.[12] Twitter also recently removed public access to its content for internet users without a Twitter account, and ended existing arrangements with Open AI to provide access to Twitter data due to insufficient compensation.[13]

4.25. Generative AI systems could also enable large digital platforms to further entrench and extend their market power by leveraging their substantial user bases and engaging in more effective and difficult-to-detect forms of anti-competitive conduct, such as anti-competitive self-preferencing and tying. In the September 2022 DPSI interim report, the ACCC recommends service-specific codes of conduct with targeted competition obligations, which would apply to designated platforms with the ability and incentive to engage in anti-competitive conduct to address such conduct.[14]

Algorithmic collusion

4.26. The use of AI algorithms also provides a way for two or more different firms to engage in anti-competitive conduct, such as in relation to setting prices, determining bids, or market sharing.[15] Collusion assisted by algorithms may make it easier for firms to avoid detection, or to effectively coordinate, where doing so may otherwise be too complicated (such as in relation to two large sets of pricing data), resulting in higher prices for customers.

4.27. One challenge for regulators is that some forms of potentially harmful algorithmic collusion are likely to be legal under current regulatory settings, including where 'competing' algorithms simultaneously learn to set higher prices collectively to maximise profit.

Media and the information environment

4.28. The ACMA is the independent statutory authority responsible for the regulation of broadcasting, and some aspects of regulation of online content delivered by digital platform services in Australia. The ACMA currently oversees the voluntary Australian Code of Practice on Disinformation and Misinformation.

4.29. Algorithms and generative AI have the potential to significantly impact the production of news and the discoverability and consumption of content and information online.

Disinformation and misinformation can be spread using AI and recommender systems

4.30. Generative AI could be used by bad actors to create and disseminate disinformation and misinformation at scale. LLMs can produce – at a very low cost – significant amounts of false information that may appear to be reliable or trustworthy. In May 2023, NewsGuard found that AI-generated sites that produced false and misleading articles, reached hundreds of thousands of followers on social media.[16]

4.31. Australians are also starting to use LLMs more regularly, and are starting to rely on LLM chatbots such as ChatGPT for authoritative answers to questions, or for

---

[12] Platformer, Reddit goes dark, 13 June 2023.

[13] Business Insider, Elon Musk cut off OpenAI's access to Twitter data, 29 April 2023.

[14] ACCC, Digital platform services inquiry – September 2022 interim report – Regulatory reform, September 2022.

[15] OECD, Algorithmic Competition, 2023, chapter 3.

[16] NewsGuard, Rise of the Newsbots: Ai-Generated News Websites Proliferating Online, 5 May 2023.

advice. As these services often generate authoritative but inaccurate responses, this can lead to Australians being given false or incomplete information.

4.32. Recommender systems that work to support user engagement on digital platforms may also contribute to the promotion of controversial, false or misleading stories, partly because these stories spread faster, and keep users engaged. In 2018, the MIT Media Lab found that false news stories spread at six times the rate of factual stories on Twitter.[17] False stories that spread quickly may include 'fringe' content that users may not have otherwise seen.

4.33. Nevertheless, algorithms can play a key role in the detection and moderation of disinformation and misinformation. While not perfect, algorithms can be employed by platforms to filter false and misleading information before it starts to spread. Platforms can support the use of algorithms for content moderation while also retaining guardrails, such as human-based content moderation, to make decisions about complex content based on local, cultural and political contexts.[18]

4.34. The voluntary *Australian Code of Practice on Disinformation and Misinformation* (the code), which is managed by the Digital Industry Group Inc (DIGI), requires signatories to provide safeguards against harms that may arise from disinformation and misinformation. It also may help to improve transparency around how recommender systems are used by platforms to address and moderate disinformation and misinformation (through outcome 1e of the code). Additionally, the Government is currently consulting on new regulatory powers to combat misinformation.

The news sector is increasingly relying on AI and recommender systems

4.35. Recommender systems are commonplace in the online news environment. While the systems can help deliver the most relevant news stories to a user based on their past behaviours, personal characteristics and interests, there can have a range of unintended negative consequences. Recommender systems may show users more sensationalist 'clickbait' articles in their news feeds, designed to elicit strong emotions and generate reactions, eroding perceptions of credibility and quality in news media.[19]

4.36. Generative AI tools draw information from a wide variety of sources, including news and media platforms. Industry stakeholders have asserted that generative AI companies should remunerate media companies for the use of their content.[20]

4.37. Generative AI is playing an increasingly important role within legitimate media organisations – supporting the creation and distribution of original journalism. While many news organisations recognise current limitations around the reliability and accuracy of these tools,[21] deploying the technology with appropriate transparency and editorial oversight may help news organisations – at lower cost – generate ideas for articles, research or interrogate large data sets, identify errors

---

[17] Dizike P, Study: On Twitter, false news travels faster than true stories, *MIT news,* 8 March 2018, accessed 3 July 2023.

[18] Caplan R, Content or Context Moderation? Artisanal, community-reliant, and industrial approaches, Data & Society Research Institute, 14 November 2018.

[19] Molyneux L and Coddington M (2020), 'Aggregation, Clickbait and Their Effect on Perceptions of Journalistic Credibility and Quality', *Journalism Practice*, 14(4):429-446.

[20] Shteyman J, 2023, 'News Corp calls for 'fair share' of AI revenue', *The Canberra Times,* 12 May 2023.

[21] See, for example, Viner K and Bateson A, The Guardian's approach to generative AI, 16 June 2023.

or suggest corrections, and reduce time spent on business processes and administration.

4.38. Existing arrangements through broadcasting codes of practice that place obligations on factual content in news and current affairs programming can also be used to hold the broadcasting industry to account for providing accurate information to audiences.

Privacy

4.39. The OAIC regulates the *Privacy Act 1988* (Cth) (the Privacy Act), which applies to the handling of personal information. Privacy obligations will apply where personal information is used to train, test or deploy algorithms within an AI system.

4.40. AI can have significant impacts on privacy. For example, the information handling practices associated with this technology are often complex and opaque which challenges the ability of individuals to meaningfully understand how their personal information is being handled. Outputs from AI systems may also contain misleading or inaccurate information about an individual.[22] In addition, the use and retention of large data sets to develop and deploy this technology increases the risk of a data breach and the risk of harm to individuals.[23]

4.41. The Discussion Paper recognises the potential for individuals' data to be used in AI in ways that raise privacy concerns. Given these concerns, strong and effective privacy protections are essential to promote the use of AI in ways that are aligned with community expectations and to foster public trust and confidence in the use of these systems.

4.42. To this end, the Privacy Act contains the Australian Privacy Principles (APPs), which apply to Australian Government agencies and private sector organisations with an annual turnover of more than $3 million, subject to some exceptions (collectively referred to as 'APP entities').[24] The APPs outline how APP entities are permitted to handle personal information and are structured to reflect the information lifecycle, from collection, through to use and disclosure, storage and destruction.

4.43. The APPs include obligations to notify individuals about the handling of their personal information, limitations on collecting personal information (including where the personal information is collected through being created by an algorithm), limitations on use and disclosure of personal information, and providing mechanisms for individuals to access and correct their personal information. This sets clear requirements for the handling of personal information.

4.44. The Privacy Act is principles-based and technology neutral, which has a number of advantages in the context of AI.

4.45. The principles-based nature of the APPs provides APP entities with the flexibility to take a risk-based approach to the protection of individuals' privacy, having regard to their particular circumstances, including size, resources and business model. This enables the APPs to be scalable and adaptable to the different acts, practices and technologies of APP entities while, importantly, allowing APP entities to

---

[22] See OAIC, Guide to data analytics and the Australian Privacy Principles, 21 March 2018, in relation to analytics processes.

[23] See, for example, Bell G, Burgess J, Thomas J., and Sadiq S, Rapid Response Information Report: Generative AI - language models (LLMs) and multimodal foundation models (MFMs), 24 March 2023; National Cyber Security Centre (UK), ChatGPT and large language models: what's the risk?,14 March 2023.

[24] *Privacy Act 1988* (Cth) s 6C and 6D (definition of 'APP entity').

simultaneously innovate and carry out their functions and activities. It also allows the Privacy Act to complement other legislation or regulatory frameworks that may deal with related issues.

4.46. The technology neutral application of the APPs enables them to apply to the handling of personal information across a diverse range of technologies, including AI. This allows for greater 'future-proofing', which preserves the relevance and applicability of the APPs in a context of continually changing and emerging technologies.[25] For example, the OAIC's Guide to data analytics and the Australian Privacy Principles provides guidance on the application of the APPs to modern data analytics despite significant advances in the generation and treatment of data since the APPs commenced nearly a decade ago. Given the 'speed of innovation in recent AI models', this future-proofing is essential to effective regulation.[26]

4.47. By contrast, detailed rules-based and technology-specific regulation is comparatively rigid. It may impose requirements that are not always appropriate for all entities regulated by the scheme, and may inadvertently result in regulatory gaps, for example, by not covering all entities intended to be regulated.

4.48. The Privacy Act contains mechanisms that allow the APPs to be supplemented by more specific rules in regulations or other legislative instruments in appropriate circumstances. For example, APP codes can adapt and particularise the APPs where appropriate, providing greater clarity on obligations where that is warranted by the entity's particular circumstances.[27]

4.49. As the Discussion Paper notes, the Attorney-General's Department's Privacy Act Review report has proposed the introduction of several new measures to enhance the current privacy regime. Box 1 of the Discussion Paper identifies proposals made in the Privacy Act Review which aim to enhance transparency and individual self-management where AI systems and algorithms are used.[28] These proposals and others are discussed in more detail in the OAIC's submission to this Discussion Paper. Many of the proposals made through the Privacy Act Review will assist to mitigate the potential privacy risks of AI systems.

Online safety

4.50. AI poses various benefits and risks to the online safety of Australians. In particular, generative AI technologies can be misused to create:

- highly realistic synthetic imagery depicting child sexual exploitation and abuse

- deepfake videos depicting individuals in sexually explicit contexts without their consent[29] or engaging in other activities that never happened

- large amounts of authentic-seeming content at scale for the purpose of bullying, abusing, or manipulating a target – including, but not limited to, grooming children for exploitation or causing people to 'pile on' a victim.

---

[25] OAIC, Australian Privacy Principles guidelines, July 2019, accessed 26 November 2020.

[26] Discussion Paper, p 3.

[27] *Privacy Act 1988* (Cth) Part IIIB.

[28] Specifically, the Discussion Paper notes proposals to enhance privacy policies by including information about whether personal information will be used in ADM which has a legal, or similarly significant effect on an individual's rights, and how APP entity may target users (including through algorithms and profiling). The Discussion Paper also notes proposals to introduce an individual right to request information about how ADM decisions are made and to opt-out of targeted advertising.

[29] Farid H., *Creating, Using, Misusing, and Detecting Deep Fakes*., Journal of Online Trust and Safety, Vol. 1, No. 4, 2022.

4.51. The *Online Safety Act 2021* (Cth) (Online Safety Act) provides eSafety with a range of regulatory functions to mitigate these and other risks.

Complaints-based investigations schemes

4.52. eSafety's four complaints-based investigations schemes capture AI-generated images, text, audio, and other content which meets the legislative definitions of:

- class 1 material (such as child sexual exploitation material and terrorist and violent extremism content) and class 2 material (such as pornography)

- intimate images produced or shared without consent (sometimes referred to as 'revenge porn')

- cyberbullying material targeted at a child

- cyber abuse material targeted at an adult.

4.53. Under these schemes, eSafety can provide support to complainants, including assisting in the removal of certain content and providing guidance to minimise the risk of further harm.

Industry regulatory schemes

4.54. The Online Safety Act also regulates online services' systems and processes through two regulatory schemes.

Basic Online Safety Expectations

4.55. eSafety can require a range of online services including social media services, messaging services and other apps and websites to report on the reasonable steps they are taking to comply with the Government's Basic Online Safety Expectations (BOSE). This is intended to enhance transparency and accountability, and to ensure people can use their services in a safe manner.

4.56. eSafety has issued 13 reporting notices since August 2022, requiring companies to report on the steps they are taking to implement the BOSE. Each notice has included questions about the use of AI tools to detect illegal and harmful content and activity, such as child sexual exploitation and abuse. A report summarising the responses from the first seven notices was published in December 2022.[30] In the future, eSafety could require other service providers to report on the reasonable steps they are taking to ensure the safety of their generative AI functionalities.

4.57. Service providers are required to respond to the notices and non-compliance with the expectations could result in a published statement of non-compliance.

Mandatory Industry Codes or Standards

4.58. In June 2023, eSafety registered five new industry codes which will take effect on 16 December 2023.[31] They require certain online service providers to take adequate steps to reduce the availability of illegal and seriously harmful online content, such as child sexual abuse and pro-terror material. AI is one of the means service providers could utilise to automatically detect known (i.e. pre-identified and verified) child sexual abuse material and pro-terror material (see social media service code).

---

[30] eSafety, Responses to transparency notices, accessed July 2023.

[31] eSafety, eSafety Commissioner makes final decision on world-first industry codes, 1 June 2023.

4.59. eSafety will determine industry standards for relevant electronic services and designated internet services, as draft codes for these services were found not to provide appropriate community safeguards. In preparing industry standards for these sections of the online industry, eSafety will consider how proposed requirements can address risks of class 1 content, including AI generated content such as deep fake child sexual abuse.

4.60. A decision to register the code for internet search engine services is yet to be made. eSafety has asked the relevant industry associations to re-draft the code to capture recently proposed changes to search engines to incorporate generative AI features, and to address the risks associated with this new technology.

Other eSafety initiatives to support responsible AI

Tech Trends and Challenges

4.61. eSafety conducts horizon scanning and engages with subject matter experts through its Tech Trends work program. This allows us to identify the online safety risks and benefits of emerging technologies, as well as the regulatory opportunities and challenges they may present. In December 2022, eSafety published a position statement on algorithms and recommender systems, and is currently drafting a forthcoming paper on generative AI, examining LLMs and multimodal models.

Safety by Design

4.62. eSafety's Safety by Design (SbD) initiative encourages industry to anticipate potential harms and implement risk-mitigating and transparency measures throughout the design, development and deployment of a product or service. This includes providing free risk assessment tools and good practice guidance to help companies build in safety features and provide positive online experiences.

4.63. SbD should be applied to all AI products and services from the earliest stages of design and throughout their lifecycle. Based on eSafety's recent expert consultations, this could include ensuring:

- generative systems are sourcing high-quality data and information which has been cleaned of illegal, exploitative, and otherwise harmful material

- policies and processes to prevent users from generating harmful content

- watermarks and detection tools are used to identify AI-generated materials

- features are evaluated to identify and mitigate risks for diverse user groups

- clear reporting mechanisms and well-defined triage and escalation processes

- system and model cards are used to promote the improvement of models and the enhancement of their understanding by regulators, researchers, and the public.

4.64. SbD is a voluntary initiative promoted by eSafety; it is not enforceable through eSafety's legislative powers. Accordingly, eSafety has somewhat limited ability to require companies to build in risk mitigation measures at the development phase when many important safety decisions are made, as its regulatory options generally only apply after a technology has been made available to Australians. Consideration should be given to the need for ex ante regulatory oversight to apply earlier in the process to ensure effective guardrails are established before technology is publicly released.

Education and Awareness Raising

4.65. eSafety's research team is developing questions on algorithmic literacy to include in its 2024 youth survey. Findings from this research will inform eSafety's ongoing online safety programs for children and young people, their parents and carers, and educators. eSafety's education programs are underpinned by the core concepts of respect, resilience, responsibility and reasoning. These concepts continue to be of relevance to AI literacy. The research will also contribute to the international evidence base on children and young people's digital literacy.

Upcoming review of the Online Safety Act

4.66. The Australian Government has announced that the Online Safety Act will be independently reviewed in the coming year, providing an opportunity to consider its suitability to address online safety issues pertaining to AI and related issues.

5. Conclusion

5.1. We trust this submission provides DISR with information about how DP-REG promotes collaboration and coordination between regulators to understand the impacts of AI and its intersection with our regulatory frameworks. We look forward to working closely with DISR and other relevant areas of the Australian government, both individually and as part of DP-REG, in developing a response that allows Australians to safely harness the benefits of this technology.

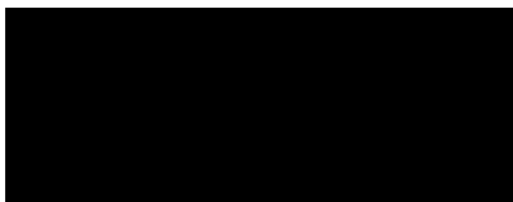5.2. We welcome further engagement with DISR in response to our submission.
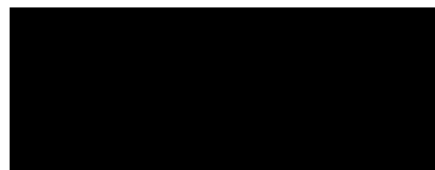
Yours sincerely

Nerida O'Loughlin PSM
Chair, Australian Communications
and Media Authority

Gina Cass-Gottlieb
Chair, Australian Competition and
Consumer Commission

Angelene Falk
Australian Information Commissioner
and Privacy Commissioner, Office of
the Australian Information
Commissioner

Julie Inman Grant
eSafety Commissioner, Office of the
eSafety Commissioner

26 July 2023