Reset.
AUSTRALIA

Response to the *Safe and Responsible AI in Australia* discussion paper

Reset.Tech Australia
July 2023

Contents

About Reset.Tech Australia & this submission

Reset.Tech Australia is an independent, non-partisan policy initiative committed to driving public policy advocacy, research, and civic engagement to strengthen our democracy within the context of technology. We are the Australian affiliate of Reset, a global initiative working to counter digital threats to democracy. We welcome the opportunity to respond to the *Safe and Responsible AI in Australia Discussion Paper*. We have framed our response around the the general proposals and overview noted in the discussion paper, as well as three specific questions:

- *Question 9: Given the importance of transparency across the AI life cycle, when and where will transparency be most critical and valuable to mitigate potential AI risks and to improve public confidence?*
- *Question 14: Do we support a risk-based approach for addressing potential AI risks?*
- *Question 20: Should a risk based approach be voluntary or self-regulated, or be mandated through regulations?*

Recommendations

- **Voluntary and co-regulatory models should be avoided** in developing the regulatory regime. They are ineffective and not harmonised with global best-practice.
- **Risk-based approaches to AI are a fruitful way forward**. However strict criteria around risk designation—including considerations around data provenance, uses and potential uses emerging from reductant capacities—needs to be drafted by regulators or legislators to avoid creating perverse incentives to inappropriately decrease risk reporting. This is in keeping with international approaches.
- Australia's approach to AI, including our risk frameworks, need to **embrace the precautionary principle**, and this needs to be reflected in any and all risk identification or mitigation measures.
- **Data provenance needs particular attention** when considering transparency across the AI lifecycle, and designating risk levels.
- **The role of consumer choice, including meaningful consent and privacy considerations**, needs to be factored in when considering risk designations and risk-mitigations.
- Particular attention needs to be given to **ensuring that any regulatory regime enshrines children and young people's best interests**.

1. Overall response to the discussion paper

Reset.Tech Australia welcomes the *Safe and Responsible AI in Australia* discussion paper ('the paper'), and its broad focus on how to manage the opportunities and challenges AI poses to Australians. The intent behind the paper, to identify potential gaps in the existing domestic governance landscape, is timely and important.

We specifically **welcome the broad and comprehensive focus on potential challenges that AI might pose**. The paper identifies eight broad challenges AI might pose Australians and the Australian economy. This includes harmful uses, but extends beyond these to also discuss inaccuracies, bias, privacy violations, at-scale risks, issues with accountability and transparency, deception and lack of consent and issues with ownership. The breadth of this focus is welcome and refreshing. Australia's current digital regulatory landscape has struggled to keep pace with the wide ranging nature of issues that technology has created, and has too often focussed on a narrow handful of online harms or limited set of privacy violations.

Relatedly, we **welcome the breadth of the regulatory landscapes identified**. The paper moves beyond the *Online Safety Act* or *Privacy Act,* to include consumer law, competition law, criminal law and so on as affected regulatory regimes. The intention to ensure that adequate tech regulation is reflected across Australia's regulatory landscape is welcome. These wide ranging policy regimes will need to be complementary, and degrees of harmonisation will be required to address any cross-cutting risks. This is no small feat, and the ambition underpinning the paper is welcome.

Noting the themes of *safety* and *responsibility*, we encourage a 'maximalist' framing of AI policy that appreciates the dynamics of corporate power and the implications from the acquisition of massive datasets by a small number of powerful companies. Incentivising safe and responsible AI will not happen through voluntary codes and declarations of corporate goodwill, it will happen through **harmonised privacy and competition law,** and an appreciation that **data policy *is* AI policy**.[1] It would be a dangerous exercise to cast AI policy responses purely into the domain of content harms. While there are indeed numerous online safety issues at the AI interface, these are downstream issues from foundational concerns of asymmetric market composition and dubious business practices.

The paper outlines a range of assumptions or suggestions that we do not agree with. Our responses to each are itemised below.

- **Voluntary guidance and co-regulatory approaches do not work**, and cannot be considered as an essential part of the regulatory mix required to maximise the opportunities from, and minimise the risks from, AI. For example, the paper discusses the potential for industry Codes, developed under the *Online Safety Act*, to address potential harms stemming from AI.[2] However, so far only one industry code has been developed under the *Online Safety Act*, and the co-regulatory process that allowed industry to draft this code ultimately led to lower safety standards (than the rest of the

[1] AI Now (2023) *Confronting Tech Power: 2023 Landscape* https://ainowinstitute.org/wp-content/uploads/2023/04/AI-Now-2023-Landscape-Report-FINAL.pdf, 12

[2] The paper presents the example of detection and algorithmic de-amplification of harmful content as one example, on page 13.

world) introduced into Australia's regulatory regime.[3] Likewise, voluntary codes such as the Misinformation and Disinformation Code have not been widely adopted.[4] (See box 1 for more details).

Co-regulation is also exceedingly unpopular with the public, who hold legitimate expectations that independent regulators will draft codes of practice. A poll of 1,502 adults in December 2022 found that only 21% of Australians trusted the social media industry to write their own codes, and the majority said they would prefer if independent regulators drafted any safety and privacy codes.[5]

Beyond demonstrable failures within broader tech issues in Australia, voluntary and co-regulation is globally being rejected as an inadequate approach to AI regulations. In the UK for example, where the Government's AI whitepaper (released in March)[6] proposed voluntary and non-binding principles, experts, technologists and civil society raised concerns. By May, the UK Prime Minister had shifted to call for a more cautious approach[7] specifically highlighting the need for regulatory guardrails at a G7 meeting.[8] The UK has since shifted far enough to appoint Ian Hogarth to chair their AI taskforce, who has previously called for precaution and oversight in the pace of AI development.[9] This brings them closer into alignment with the European approach and the global consensus. As Doreen Bogdan-Martin, the Secretary-General of the International Telecommunication Union simply put it "Business, alone, can't be self-regulating."[10] Where governments have considered it, they are inevitably rejecting it as inadequate.

Box 1: Issues with Self and Co Regulation in existing Australian digital regulations

The shortcomings of self- and co-regulation models

On regulatory models, we note that much of the regulation of digital platforms in Australia has rested on self- or co-regulatory mechanisms, from the self-regulatory disinformation and misinformation code to co-regulatory online safety codes. These mechanisms have demonstrably failed, for the following reasons:

[3] Reset Tech (2022) *How outdated approaches to regulation harm children* https://au.reset.tech/news/how-outdated-approaches-to-regulation-harm-children-and-young-people-and-why-australia-urgently-needs-to-pivot/

[4] See for example Andrea Carson (2023) 'More stick, less carrot: Australia's new approach to tackling fake news' *ASPI: The Strategist* www.aspistrategist.org.au/more-stick-less-carrot-australias-new-approach-to-tackling-fake-news/ or Jennifer Doggitt (2023) 'Review of digital industry code on misinformation fails to address public health concerns' *Croakey Health Media* www.croakey.org/review-of-digital-industry-code-on-misinformation-fails-to-address-public-health-concerns/

[5] Reset Tech (2022) *How outdated approaches to regulation harm children* https://au.reset.tech/news/how-outdated-approaches-to-regulation-harm-children-and-young-people-and-why-australia-urgently-needs-to-pivot/

[6] UK Government 2023 *AI Regulation: A pro-innovation approach* https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach

[7] Henry Zeffman 2023 'We need guardrails to regulate AI, Rishi Sunak says at G7' *The Times* https://www.thetimes.co.uk/article/we-need-guardrails-to-regulate-ai-rishi-sunak-says-at-g7-summit-97xbmbfz2

[8] Alex Wickham 2023 'Rishi Sunak Strikes More Cautious Tone on AI in Call for G-7 Regulation' *Bloomberg* https://www.bloomberg.com/news/articles/2023-05-18/rishi-sunak-strikes-more-cautious-tone-on-ai-in-uk-call-for-g-7-regulation#xj4y7vzkg

[9] Ian Hogarth 2023 'We must slow down the Race to God Like AI' *Financial Times* https://www.ft.com/content/03895dc4-a3b7-481e-95cc-336a524f2ac2

[10] Politico 2023 *Daily Bridge July 6th 2023* https://www.politico.eu/newsletter/digital-bridge/i-have-a-plan-to-fix-social-media/

- *Fails to deliver effective improvements for Australians*. For example, the draft Online Safety Codes, which the eSafety Commissioner has asked to be revised and resubmitted, demonstrate how co-regulation failed to deliver online safety and privacy for children and young people. The eSafety Commissioner initially rejected the Codes developed by industry, because they did not provide adequate community safeguards.[11] The industry drafted Codes proposed safety standards that were lower than existing online safety norms, and were below international best practice.[12] Ultimately 2 were permanently rejected, and 6 accepted on the second try. This was not an ideal outcome in terms of improving user safety nor consumer trust. Additionally, the extensive ping pong process of co-regulation itself consumed considerable time and energy from regulators, civil society and industry, without contributing to preventing harms in the interim, or even in the long term.

- *Fails to comprehensively generate necessary reforms*. The Disinformation and Misinformation Code of Practice, which is voluntary, allows platforms to in effect decide if they want to improve or not. To date, only eight platforms have signed up,[13] and many digital platforms that have known issues of mis and disinformation uncovered. For example BitChute, Odyssey and Telegram are not signatories despite being available in Australia and known vectors of disinformation and misinformation.[14] Voluntary regulation means that platforms that knowingly engage in harmful practices can simply choose to avoid improvements.

Both of these examples have ultimately required regulators to step in and take action. In the case of Online Safety, the eSafety Commissioner rejected the Codes and asked for improvements. The Commissioner are now drafting two themselves. Likewise, the shortcomings of the Misinformation and Disinformation Code process meant that ACMA now has to be given additional powers around information gathering and record keeping to enable adequate cross-platform oversight, and options to develop regulator drafted Standards as a 'fall back'.[15] If the voluntary Code were truly effective and driving improvements at scale and at pace, these measures would not be necessary.

In the EU too, obligations under the self-regulatory *Code of Practice on Disinformation* (2018) were found to be inadequate, and replaced by statutory obligations within the *Digital Services Act*. All our key allies – Canada, UK, EU – have realised that voluntary and co-regulatory approaches do not work for digital regulation and have pivoted away from them. Harm happens while we wait for self- and co-regulation to fail, and regulator-drafted industry standards should be the norm.

---

[11] Office of the eSafety Commissioner (2023) *Online industry asked to address eSafety's Concerns with the Safety Codes*
https://www.esafety.gov.au/newsroom/media-releases/online-industry-asked-address-esafetys-concerns-draft-codes-
[12] Reset Tech (2022) *How outdated approaches to regulation harm children*
https://au.reset.tech/news/how-outdated-approaches-to-regulation-harm-children-and-young-people-and-why-australia-urgently-needs-to-pivot/
[13] Adobe, Apple, Google, Meta, Microsoft, Redbubble, TikTok and Twitter. See ACMA (2022) A*ustralian Code of Practice for Disinformation and Misinformation*
https://www.acma.gov.au/online-misinformation#:~:text=you%20have%20concerns.-,Australian%20Code%20of%20Practice%20for%20Disinformation%20and%20Misinformation,%2C%20Redbubble%2C%20TikTok%20and%20Twitter.
[14] Mark Scott (2022) 'Fringe social media networks sidestep online content rules' *Politico*
https://www.politico.eu/article/fringe-social-media-telegram-extremism-far-right/
[15] Minister for Communications (2023) *New ACMA Powers*
https://minister.infrastructure.gov.au/rowland/media-release/new-acma-powers-combat-harmful-online-misinformation-and-disinformation

We note the *Privacy Act Review*[16] suggests moving towards a regulator-drafted direction for a number of Privacy Codes, which is a positive development. This should be welcomed by all who care about the privacy protections in Australia as well as democratic oversight of technology. Furthermore, it demonstrates that Australia is capable of such an approach and this can be replicated in other streams of digital regulation.

- **Regulation is pro-innovation**. In a number of places, the paper suggests regulations stifle or inhibit innovation. However, regulation and innovation can and do go hand in hand (as the paper also suggests at one point).[17] In an analysis about how the UK Government can best achieve its ambition to make the UK an AI superpower,[18] the well respected Ada Lovelace Institute prepared a comprehensive analysis of the needs of the UK's nascent AI sector for growth, and the needs of the public to trust and embrace AI. The Institute concluded that:

  "Far from being an impediment to innovation, effective, future-proof regulation will provide companies and developers with the space to experiment and take risks without being hampered by concerns about legal, reputational or ethical exposure. Regulation is also necessary to give the public the confidence to embrace AI technologies, and to ensure continued access to foreign markets."[19]

  Australian industry is no different. In order to embrace the potential of AI and for our AI industry to grow, clear, unambiguous regulation that is harmonised both internally and with major markets such as the EU is necessary. Businesses and start-ups value regulatory predictability and harmonisation. By legislating clear guardrails now, Australian companies will have certainty around what to expect and can innovate and develop accordingly.

  Further, the assumption that self/co-regulation models do not pose a burden on industry—so are therefore more pro-innovation—is misguided. Recent Australian experience also suggests this is flawed in practice. Self and co-regulatory processes place significant burdens on industry to step into the world of policy-making, reviewing legal requirements, interfacing with regulators, undertaking community consultations, including with vulnerable communities, and drafting nuanced public policy. As the recent experience of the Online Safety Codes outlined, co-regulation is not a trivial pursuit and required significant time and expertise, especially where industry misstepped and had to repeat the process to correct. (Civil society also shared this burden, with frequent and ongoing requests to engage in the process). Policy making is a unique skill set, and it places an additional burden on industry to assume that they either have or can recruit for these skills. This is especially true for small and medium industry actors, who end up being reliant on large international platforms to

[16] Attorney General's Department, (2023) *Privacy Act Review Report* https://www.ag.gov.au/rights-and-protections/publications/privacy-act-review-report
[17] Which the paper also acknowledges on page 29
[18] Department for Science, Innovation and Technology 2023 *Science and Technology Framework* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1140217/uk-science-technology-framework.pdf
[19] AdaLovelace Institute 2021 *Regulate to Innovate* https://www.adalovelaceinstitute.org/wp-content/uploads/2021/12/Regulate-to-innovate-Ada-report.pdf

lead the process that ultimately shapes the ecosystem they develop within. This burden and reliance does not help Australian innovation take root.

There are also three issues that are not addressed within the paper, presumably for the sake of brevity. Reset.Tech wishes to raise them here to ensure they are not overlooked:

- The role of precaution, and **embracing the precautionary principle** when it comes to framing our regulatory landscape. The precautionary principle has a long history of underpinning scientific and technological regulations around the world, and is referenced in AI declarations from the G7, OECD and G20. It has widespread global support. Further, precautionary frameworks—such as the UK's Inter-Departmental Liaison Group on Risk Assessment—are pro-innovation and could be effective in considering the ways to address challenges and grow opportunities for AI.[20] This includes regulation to ensure viable termination obligations are in place where effective human control over a system is no longer possible.

- The **role of consumer choice** when it comes to managing the challenges of AI. Where consumers are affected by AI—from recommender algorithms, to ADM—opt-outs are important and opt-ins are often more powerful. We specifically highlight these because they are being contested with the realm of the *Privacy Act* reforms, where the AdTech industry appears to be mounting a defence against providing consumers with an opt-out for targeted advertising.[21] Providing consumers with meaningful choice and the ability to consent or decline from the use of AI involving their data is important to engender trust.

- We would also like to restate the importance of **ensuring that any regulatory regime pays particular attention to the rights of children and young people**. Children and young people are uniquely vulnerable to AI harms (see box 2), and their best interests need to be considered and advanced in any regulatory regimes. The introduction of the 'best interests' principle into AI governance would be in keeping with emerging international norms, and harmonise with proposals put forward for privacy protections in the *Privacy Act* review.

Box 2: The risks of ADM to children and young people's rights

ADM can potentially be used to shape and limit future opportunities for children and young people[22] in ways that are concerning.

Young people are particularly vulnerable to ADM because it involves the use of personal data to make what can be extremely significant decisions that shape life chances. Unlike previous generations, today's young people are datafied from before birth[23] and will carry around a whole lifetime's worth of data. The vastness of their data footprint means that

[20] See for example, the Carnegie Trust's submission regarding the UK's AI Strategy. Carnegie Trust (2023) *Carnegie UK Response to AI Strategy White Paper* https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2023/06/28131108/Carnegie-UK-Response-to-AI-Strategy-White-Paper.pdf

[21] Sam Buckinham-Jones 2023 'Privacy overhaul 'goes beyond any other any country' *AFR* https://www.afr.com/companies/media-and-marketing/privacy-reforms-undermine-businesses-using-instagram-facebook-meta-20230517-p5d92k

[22] Referring to people under the age of 18

[23] Veronica Barassi 2020 *Child | Data | Citizen* MIT Press

events in their past could continue to shape their future in ways we cannot even begin to imagine. Data about a heart murmur detected in a baby wearable, or a mental health crisis at 13, could be used to deny access to medical insurance in your 60s. A whole childhood's worth of data could be used to determine or limit eligibility for future opportunities. This risk is not in their best interests.

Aside from damaging future opportunities, here-and-now data can also be used to shape childhoods in ways that aren't fair or amplify existing bias. Data and algorithms are already used to determine a range of childhood experiences in Australia. For example, young people completing school in NSW are awarded a HSC score which is automatically converted into an ATAR score for university entry by an algorithm[24]. Decisions around child support payments in Australia can be legally made by computer programmes. [25] While both these systems may be functioning perfectly and there have been no accusations of bias in them, there are international examples where these exact systems have failed children in unfair and biassed ways. In 2020 in the UK, a 'mutant algorithm' converted high school grades into university entrance criteria, in a way that systematically downgraded the scores of children from low-income areas, harming low-income children[26] This caused a significant public backlash, and the Government of the day had to U turn and abandon the use of algorithmically calculated grades, further eroding public trust. In 2021, in the Netherlands, data about children's ethnicity was used to automatically 'red-flag' child support payments, leading to Black and dual national families automatically having payments stopped in error, harming family incomes and is associated with more than 2,000 children being taken into care.[27] This too, caused public outroar and led to the resignation of the Prime Minister and the entire Cabinet of the time, deeply harming public trust. And in Australia too, many young people and families with children were caught up in the Robodebt scandal, often with disastrous consequences.[28] These risks are demonstrably not in children's best interests.

[24] UAC 2015 Calculating the Australian Tertiary Rank in New South Wales https://www.uac.edu.au/assets/documents/atar/atar-technicalreport.pdf

[25] Child Support (Assessment) Act 1989 (Cth) Section 12A https://www.legislation.gov.au/Details/C2016C00954

[26] Sean Coughlan 2020 'A Levels and GCSEs: Boris Johnson blames mutant algorithm for exam fiasco' *BBC* https://www.bbc.com/news/ education-53923279

[27] SeeeEuropean Parliament Anti Racism & Diversity Group 2021 'Condemning the Dutch Child Benefit Scandal' https://www.europarl.europa.eu/doceo/document/O-9-2022-000028_EN.html and NL Times (2022) 'Far More Children Taken from Homes of Victims in Tax Scandal' *NL Times* https://nltimes.nl/2022/11/30/far-children-taken-homes-victims-tax-office-benefits-scandal

[28] Luke Henriques-Gomes 2020 'Robodebt official challenged by mothers of two young men who took their own lives' *The Guardian* https://www.theguardian.com/australia-news/2020/aug/17/robodebt-official-challenged-by-mothers-of-two-young-men-who-took-their-own-lives

2. Response to specific questions

Question 9a&b

A. Given the importance of transparency across the AI life cycle, when and where will transparency be most critical and valuable to mitigate potential AI risks and to improve public confidence?

AI transparency should be understood more broadly than just explaining the technology that makes it function, to—as the paper suggests—encompass broader aspects like system risks and limitations, attributing responsibility or liability, as well as disclosing the social, cultural, and organisational context of its use.

Transparency requirements for AI systems developers should be framed around continuous risk and impact assessment in relation to the AI's design goals, with regular reporting undertaken throughout the system's lifecycle. Organisations should be required to make clear both *when* and *how* they are developing and using AI systems either to the relevant regulator(s), dedicated oversight body and/or directly to the end user depending on its use and maturity..

Transparency reporting requirements might include:

- **Procedural transparency**: how AI algorithms are built and overseen
- **Content transparency**: the provenance of data used by algorithms, how it is collected and processed in accordance with the *Privacy Act* (including any updates);
- **Outcome-based transparency**: how algorithms, the data (including training data) they use and their decision-making outputs are iterated, fine-tuned, evaluated and implemented. Known unknowns about the AI systems' unpredictability and future functionality in data processing and decision-making should be clearly indicated; impact assessments detailing how individuals and the wider ecosystem might be affected with consistent measures of effectiveness vs harm.

As this suggests, transparency is critical across the whole AI life cycle to ensure trust and safe use. However, we would like to emphasise the importance of **trust, privacy and consent in data providence,** as issues that are frequently overlooked.

Data is often needed en masse to train LLMs, ADMs and MfMs. This raises issues around privacy and consent, and fairness and reasonableness where data is used in ways that the public might not expect. For example, children's data is often swept up and used in training AI. While many models then place guardrails and restrictions on their use regarding children (i.e. they might not be able to be used to generate content that mimics or related to children), this does not undue the violations associated with the unfair and unreasonable collection and use of children's data in the first instances. (Often also without adequate parental consent or child's assent). Transparency about training sets is critical to both improve public trust and prevent future harms resulting from unanticipated redundant capacities.

Where AI models have been trained using data that cannot demonstrate consent nor pass a fair and reasonable use test, they may be designated unacceptable or high risk models by default.

Likewise, **trust in content (output) providence** is also important for generative AI products. Being able to understand and track what is synthetic content and what is not, in robust and meaningful ways, will also help ensure public trust in our information architecture.

B. Given the importance of transparency across the AI life cycle, how should transparency requirements be mandated across the public and private sectors?

Transparency requirements must be applied equally to the public and private sectors. This is both because the potential for harm stems from both sectors, and because the complex chain of tech integration creates interdependencies that cannot meaningfully be separated. For example, private sector entities often undertake public sector work using AI, and public sector data often trains private sector AI technology. Given the nature of Australia's AI landscape, regulatory regimes for the public and private sector must be equally transparent. Meaningful transparency does not need to undermine IP nor share trade secrets, so private sector exemptions are not necessary.

Question 14

Do we support a risk-based approach for addressing potential AI risks?

Yes. Risk-based approaches have proved to be flexible and (relatively) platform/company neutral and future proof in other digital regulatory spaces. They are also embraced by a number of other countries and regions, notably the EU and UK, which should reduce regulatory friction for Australian industries wishing to export. They are particularly effective when the precautionary principle is embraced.

However, to be effective and sufficiently clear for technologists to use easily, the criteria for what counts as low-, medium-, and high-risk needs to be clear, transparent, and subject to regulatory review. **Where companies can self-designate without regulatory oversight and intervention, it creates a perverse incentive to designate technologies as lower risk.** Regulators and legislators must draft the risk assessment criteria.

Risk assessments and frameworks should be designed to ensure that

- Some uses and potential uses of AI—including AI that uses data with questionable data providence—must always be designated high risk as a precautionary measure.
- There should also be an additional level within the framework to categories "unacceptable risks" that should not be accepted under any circumstances. This is in keeping with emerging European responses.
- Requirements must be in place to ensure terminal obligations where human control over systems ceases to be possible.

The template risk assessment provided in the paper is a good start, however we would recommend that the need for regulator or peer review is extended to high and potentially medium risk uses rather than just very high risk uses.

Question 20

Should a risk based approach be voluntary or self-regulated, or be mandated through regulation?

As discussed above:

- **Self and co-regulation has proven ineffective** in ensuring trust, safety or privacy in digital technologies. Voluntary and co-regulatory Codes in both the safety and mis and disinformation space have delivered a regulatory regime that is:
    - Less protective than comparative international norms, and
    - No more protective than having no regulation.
- Many of the companies involved in these processes are also leaders in the use of AI, and there is no evidence to suggest that repeating this approach would lead to different outcomes in AI regulation.
- Many other countries and regions are actively rejecting voluntary and co-regulatory models for regulating AI.
- Self and co-regulation also fails to generate public trust, as polling has shown.