CYBER SECURITY
COOPERATIVE
RESEARCH
CENTRE

SUBMISSION: Safe and Responsible AI in Australia – Discussion Paper

cybersecuritycrc.org.au

To whom it may concern,

Submission: *Safe and Responsible AI in Australia – Discussion Paper*

I am pleased to submit the Cyber Security Cooperative Research Centre's (CSCRC) response to the consultation on the *Safe and Responsible AI in Australia – Discussion Paper* (the Discussion Paper) by the Department of Industry, Science and Resources.

The Discussion Paper is an important first step defining vital regulatory oversight of artificial intelligence (AI) technologies in Australia. These diverse technologies are proliferating rapidly and, while they can and do bring efficiencies, they can also have a deleterious effect on human decision making. Therefore, a balance needs to be struck between innovation, regulation and human oversight of AI. In our view, this is highly pertinent in relation to the intersection between AI and cyber security, where the application of AI technologies offers significant benefits but also has to potential to empower malicious actors.

This submission does not address every question raised in the Discussion Paper, with responses supplied where we believe we can meaningfully contribute.

About the CSCRC

The CSCRC is dedicated to fostering the next generation of Australian cyber security talent, developing innovative projects to strengthen our nation's cyber security capabilities. We build effective collaborations between industry, government and researchers, creating real-world solutions for pressing cyber-related problems.

By identifying, funding and supporting research projects that build Australia's cyber security capacity, strengthen the cyber security of Australian businesses and address policy and legislative issues across the cyber spectrum, the CSCRC is a key player in the nation's cyber ecosystem.

We look forward to answering any queries about this submission and welcome the opportunity to participate in any future consultation.

Yours Sincerely,

Rachael Falk

CEO, Cyber Security Cooperative Research Centre
ceo@cybersecuritycrc.org.au

Introduction

Artificial intelligence (AI) technologies have the potential to be transformative, bringing new efficiencies to the way we work, live and do business. But with the good comes the bad and, from a cyber security perspective, while AI can serve to bolster defences it can also be used by malicious actors to enable more sophisticated exploits and attacks. For this reason, the Cyber Security Cooperative Research Centre (CSCRC) submits that in developing a plan for the safe and responsible use of AI in Australia, cyber security must be a primary concern.

In producing this submission, the CSCRC asked Generative AI platform ChatGPT several questions about the AI regulation, its impacts on cyber security and the ethical application of AI technologies. These are excerpts from ChatGPT's responses:

"The question of whether AI should be regulated in Australia, or any other country, is a complex and debated issue. The regulation of AI involves balancing the potential benefits and risks associated with its development and deployment."

"AI can both enhance and pose challenges to cybersecurity. It has the potential to significantly improve cybersecurity defenses, but it can also be exploited by malicious actors … Given these challenges, it is crucial to develop robust cybersecurity strategies that address the potential risks associated with AI. This includes implementing AI-powered defense mechanisms, ensuring the security of AI systems themselves, and promoting research into adversarial attacks and defenses."

"AI itself is not inherently dangerous. It is a tool or technology that is created and controlled by humans. The potential dangers associated with AI stem from how it is developed, deployed, and used. AI systems can be powerful and complex, and their impact depends on various factors such as the quality of the data they are trained on, the algorithms used, and the decisions made by their human creators."

These responses, produced in a matter of seconds, speak to the sophistication of Generative AI and the power it yields. However, these technologies do not just answer questions – they can produce malicious code, sophisticated fraudulent content and, in the context of deep fakes, can create life-like audio, imagery and visual recordings that can be used for nefarious purposes. Therefore, as recently noted by Lindy Cameron of the UK's National Cyber Security Centre (NCSC): "The scale and complexity of these models is such

that if we don't apply the right basic principles as they are being developed in the early stages it will be much more difficult to retrofit security"[1].

Key to the effective regulation of AI is enhancing public understanding of what AI technologies are, noting their diversity in form and application and how they are used across a range of areas, including in the everyday lives of citizens. For this to occur, transparency and AI literacy are key – and public understanding of AI in Australia is lagging. University of Queensland research indicates most Australians (61%) have a low understanding of AI, including how and when it is used. For example, while 78% of Australians report using social media, 59% were not aware social media applications use AI[2]. Overall, this research found that public trust in AI systems would be improved by strengthening the regulatory framework for governing AI and strengthening Australia's AI literacy[3].

Furthermore, Australia's recent experience of automated decision-making gone wrong, laid bare by the *Royal Commission into the Robodebt Scheme* and its recommendations[4], should serve as a stark wake up call to legislators and policy makers that integrity and security should be 'baked in' to the use of emerging technologies. Because poor governance can result in serious adverse outcomes for ordinary people. As highlighted by CSCRC CEO Rachael Falk recently in the Sydney Morning Herald: "Ultimately, AI is only as good as the algorithm that operates it, the data that trains it and the law that underpins it"[5].

When it comes to the regulation of emerging technologies, there has been a trend domestically and globally of regulating too late and trying to play catch up – cyber security is a prime example. In relation to AI technologies, which can unlock significant potential but also result in serious harms, swift regulatory action must be a key priority for the Federal Government. Not only will this help build public trust in AI technologies, it will create guardrails and certainty for organisations investing billions in AI development and implementation and ensure Australia keeps pace with the rest of the world. And, most importantly, human oversight must form a cornerstone of regulation, establishing checks and balances to ensure that any risks, harms or bias are detected and effectively addressed.

[1] AI must have better security, says top cyber official - BBC News
[2] Australians have low trust in artificial intelligence and want it to be better regulated (uq.edu.au)
[3] Ibid 2
[4] Royal Commission into the Robodebt Scheme
[5] Robo-debt is possible again if we give AI too much power (smh.com.au)

Definitions:

1. Do you agree with the definitions in this discussion paper? If not, what definitions do you prefer and why?

Yes, the definitions included in the discussion paper are currently fit for purpose. However, given the rapid pace of AI development and new and emergent forms of AI, these definitions should be reviewed annually to ensure they remain suitable and capture all necessary components. Such review should occur concurrently with any review of the definitions included in Australia's *List of Critical Technologies in the National Interest*[6].

Potential gaps in approaches:

2. What potential risks from AI are not covered by Australia's existing regulatory approaches? Do you have any suggestions for possible regulatory action to mitigate these risks?

In relation to potential AI risks that are not covered by existing regulatory approaches, there are two key areas of concern for the CSCRC. These are:

- Risk of serious harms resulting from the use of AI; and
- Cyber security implications of AI

Harms

How to effectively deal with serious harms resulting from digitally-enabled technologies has been an ongoing issue for Australian policy makers, with much of the debate centred around privacy. The Australian Law Reform Commission's (ALRC) 2014 report *Serious Invasions of Privacy in the Digital Age* recommended "that a statutory cause of action for serious invasions of privacy should be contained in a new, stand-alone Commonwealth Act"[7]. This is an issue that has been revisited as part of the *Privacy Act Review Report*, released in late 2022, which recommends that a statutory tort for serious invasions of privacy be introduced (Proposal 27.1). The report also proposes that, in relation to automated decision making, a right for individuals to request meaningful information about how substantially automated decisions with legal or similarly significant effect are made be introduced (Proposal 19.3)[8]. Given *Privacy Act 1988 (Cth)* (Privacy Act) reforms are still under consideration, there is scope to harmonise any AI-specific regulation with proposed amendments to the Privacy Act and, more specifically, ensure AI technologies are captured

[6] List of Critical Technologies in the National Interest | Department of Industry, Science and Resources
[7] A New Tort in a New Commonwealth Act.docx (alrc.gov.au)
[8] Privacy Act Review Report 2022 (ag.gov.au)

within the Privacy Act's definitions (noting proposed revised definitions within the Privacy Act do seek to capture a range of technologies, including AI).

In relation to AI harms, and as noted in the introduction of this submission, there are key learnings to be taken from the *Royal Commission into the Robodebt Scheme*. Furthermore, there is a mandate. As part of her wide-ranging findings, Commissioner Catherine Holmes made several key recommendations related to data-matching and exchanges and automated decision-making. These practical recommendations include the establishment of an oversight body to monitor and audit automated decision-making and a strengthening of the governance related to data-matching programs[9].

Finally, the issue of AI selection bias risks must be considered in relation to potential harms arising from these technologies. Because of their reliance on data and training to work, AI technologies can – and have - produced outputs that are biased or incorrect, resulting in harms. There have been key examples of such selection bias occurring in the US in relation to predictive policing, through which minority groups have been targeted[10], and in relation to judicial decision making, with AI programs used to predict the potential of a recidivism amongst convicted criminals[11].

Cyber security implications

AI technologies – notably Generative AI – have the potential to create new and sophisticated cyber security threats. Some of these risks were summarised well in a recent Forbes article, which notes that "AI-generated outputs can … be used to make cybercrime more lucrative and convincing—whether it's launching a social engineering attack, fine-tuning malware code to make it harder to detect or using AI to generate and share guidelines, advice and tutorials with cybercriminals"[12]. While several regulatory frameworks do infer that organisations must prepare for current and emerging cyber risks, notably the *Security of Critical Infrastructure Act (2018)* (SOCI Act) and the Australian Prudential Regulatory Authority's (APRA) prudential standards *CPS 234 – Information Security*, there is a lack of prescriptive guidance across Australia's broader regulatory ecosystem as to how cyber risks should be managed in the context of new and emerging technologies. This is an issue that could potentially be explored if the Department of Home Affairs is to pursue the establishment of a domestic Cyber Security Act. Furthermore, there is scope for government to better clarify how disruptive technologies like AI will impact

[9] Report | Royal Commission into the Robodebt Scheme
[10] Predictive policing algorithms are racist. They need to be dismantled. | MIT Technology Review
[11] Machine Bias — ProPublica
[12] Preparing For The Cyber Risks Of Generative AI (forbes.com)

organisational cyber resilience and how boards should manage new and emerging threat vectors in the context of directors' duties as defined in the *Corporations Act 2001*.

3. Are there further non-regulatory initiatives the Australian Government could implement to support responsible AI practices in Australia? Please describe these and their benefits or impacts.

In the CSCRC's view, there are three key areas of non-regulatory action the government can address to help build public and private sector understanding of AI and its relationship to cyber security, as well as supporting enhanced knowledge in the general community. These include:

- Public education;
- Sector-specific guidance; and
- Research guidance.

Public education: Australians lack a strong understanding of AI. University of Queensland research indicates most Australians (61%) have a low understanding of AI, including how and when it is used. This research highlights, for example, that while 78% of Australians reported using social media, 59% were not aware social media applications use AI[13]. However, public education can support citizens become more AI literate, as has been evidenced in Finland, where the government has partnered with the private sector to deliver a free AI literacy course[14]. The course, *Elements of AI*, aims to demystify AI, raising awareness about the opportunities and risks of AI among members of the general public[15]. The CSCRC also submits that review of the Australian Curriculum's 'Technologies' learning area could be undertaken to incorporate education about emerging technologies, such as AI.

Sector-specific guidance: There are specific sectors and groups that would benefit from more detailed guidance as to the use of AI. The CSCRC notes the government has released guidance related to the use of *Generative AI by Australian Public Service*[16], which could be used as a blueprint for general guidance by regulators at a sectoral level. There is also scope for the government to provide guidance for increased clarity as to how private

[13] Ibid 2
[14] Finland's Visionary Initiative: AI Literacy | by Stylumia | Stylumia | Medium
[15] A free online introduction to artificial intelligence for non-experts (elementsofai.com)
[16] Interim guidance for agencies on government use of generative AI platforms | aga (digital.gov.au)

entities and company directors should address governance issues pertaining to the use of AI in the private sector.

Research guidance: For researchers, there is scope to for government to provide guidance as to the use of AI in research undertaken in Australia. While the ethics of AI use in research has been a topic of much discussion globally, little attention has been paid to the security of research data and protection of intellectual property (IP). In relation to IP, while issues surrounding copyright of images, music and patent violations have been a topic of significant discussion, the protection of research IP has not gained prominence. This is an issue that also needs to be considered by university peak bodies and individual universities in relation to the protection and integrity of IP produced by academics.

4. Do you have suggestions on coordination of AI governance across government? Please outline the goals that any coordination mechanisms could achieve and how they could influence the uptake of AI in Australia.

The CSCRC submits there is scope for the Federal Government to establish a stand-alone body to coordinate AI governance across government and the economy more broadly. For example, to help guide and enforce regulation of Australia's SOCI Act regime, the Cyber and Infrastructure Centre (CISC) was established within the Department of Home Affairs in 2022. Despite being operational for just a short time, the CISC has established itself as a trusted and proactive force in driving and supporting SOCI Act implementation, taking a truly collaborative approach. Hence, we submit duplicating such model to create an Emerging Technologies Regulatory Centre within the Department of Industry, Science and Resources (DISR), or another suitable department, should be considered.

Similarly, in response to a spate of serious and high-profile cyber incidents impacting Australian-based organisations, the role of National Cyber Security Coordinator was established within the Department of Home Affairs. The Coordinator is tasked with leading national cyber security policy, the coordination of responses to major cyber incidents, whole-of-government cyber incident preparedness efforts and the strengthening of Commonwealth cyber security capability.[17] Following this lead, the CSCRC submits that within DISR, or another suitable department, there is scope to appoint an Emerging Technologies Coordinator. The Coordinator could actively support legislative and regulatory responses to the development and adoption of emerging technologies in Australia, including AI. Likewise, the Coordinator could play a key role in liaising between government

[17] Cyber Coordinator (homeaffairs.gov.au)

and industry, helping ensure the development and implementation of emerging technologies occurs safely and ethically without hindering innovation.

Responses suitable for Australia:

5. Are there any governance measures being taken or considered by other countries (including not discussed in this paper) that are relevant, adaptable and desirable for Australia?

In the CSCRC's view, the European Union's AI Act, which is set to be enacted later this year, sets a precedent for others to follow. As noted by experts at the Carnegie Endowment for International Peace, the AI Act "strikes a balance between the dual imperatives of providing predictability and keeping pace with AI developments. Its risk-based approach allows regulators to slot new application areas into existing risk categories as AI's uses evolve, providing a balance between flexibility and regulatory certainty. Meanwhile, the AI Act's definition of relatively flexible essential requirements also alleviates the key precision challenge posed by purely horizontal frameworks, allowing compliance strategies to be flexible across sectors and as technology evolves".[18]

While the AI Act is not perfect and, upon its enactment, various regulatory and technical issues will undoubtedly have to be addressed, it provides a solid blueprint for what an Australian legislative approach to regulating AI could look like. Furthermore, given Australia's strong trade and cultural ties with the EU, the adoption of a similar regime would support cross-border interoperability and alignment.

Target areas:

7. How can the Australian Government further support responsible AI practices in its own agencies?

Firstly, and as noted in the response to Question Two, the Federal Government should act quickly to implement relevant recommendations from the *Royal Commission into the Robodebt Scheme*. Central to enhancing such responses will be the need for human oversight and appropriate manual checks and balances to help prevent a situation like Robodebt occurring again. This will involve a whole-of-government approach, drawing on the expertise from across government, including legal guidance from the Attorney-General's Department, to establish solid foundations for the application of AI across departments and agencies.

[18] Lessons From the World's Two Experiments in AI Governance - Carnegie Endowment for International Peace

Furthermore, the Federal Government could consider enacting a directive like the Canadian Government's *Directive on Automated Decision-Making* (the Directive). The aim of the Directive is to "ensure that automated decision systems are deployed in a manner that reduces risks to clients, federal institutions and Canadian society, and leads to more efficient, accurate, consistent and interpretable decisions made pursuant to Canadian law"[19]. Under the Directive data and information on the use of automated decision systems in federal institutions are made available to the public, where appropriate, Algorithmic Impact Assessments must be conducted prior to the production of any automated decision system and be published with public access.

Finally, consideration should be given to introducing a requirement for all government departments and agencies to produce annual transparency reports regarding their use of AI, including disclosures of any significant issues that have arisen as a result of the application of AI technologies. Such information could be incorporated into annual reporting requirements to reduce duplication.

9. Given the importance of transparency across the AI lifecycle, please share your thoughts on:

b. mandating transparency requirements across the private and public sectors and how these requirements could be implemented.

As noted above, the CSCRC submits that government departments and agencies should produce annual transparency reports regarding their use of AI, including disclosures of any significant issues that have arisen as a result of the application of AI technologies. Such information could also be incorporated into annual reporting requirements.

In relation to the private sector, the CSCRC submits that a threshold limit for transparency reporting could be applied, based on an organisation's annual turnover. Such a regime could be based on other frameworks with threshold limits, for example, the *Modern Slavery Act 2018*, which requires large businesses and other entities in the Australian market with annual consolidated revenue of at least A$100 million to prepare annual Modern Slavery Statements[20].

[19] Directive on Automated Decision-Making- Canada.ca
[20] Modern slavery | Attorney-General's Department (ag.gov.au)

Implications and infrastructure:

13. What changes (if any) to Australian conformity infrastructure might be required to support assurance processes to mitigate against potential AI risks?

Standard setting will be vital to mitigating against potential AI risks, while also ensuring harmonisation and interoperability with other international regimes. In relation to AI, standard setting is an emerging field and therefore, observing and aligning to international best practice in this space. Hence, it will be important to analyse the effectiveness of the EU's standard setting in relation to the AI Act and its implementation and to also consider the adoption of AI-related National Institute of Standards and Technology (NIST) standards in Australia.

Risk-based approaches:

14. Do you support a risk-based approach for addressing potential AI risks? If not, is there a better approach?

Yes – the CSCRC supports a risk-based approach, underpinned by clear regulation and guidance, for addressing potential AI risks.

15. What do you see as the main benefits or limitations of a risk-based approach? How can any limitations be overcome?

The main benefit of taking a risk-based approach to AI regulation is that individual organisations can assess potential risks in a way that relates to their unique operations, prioritising risk management in line with the most substantial threats to their business continuity. Such an approach also encourages organisational agency and the ability to pivot quickly in the event new risks arise, or the threat environment changes.

Risk-based approaches are not without challenges and some organisations have called for more prescriptive approaches to regulation. This occurred in relation to the SOCI Act, which takes an all-hazards risk-based approach to managing threats to critical infrastructure. During the SOCI Act consultation, several organisations raised concerns about the legislation being overly broad, making implementation difficult[21]. However, since the regime has come into force, informative and detailed guidance released by the CISC has supported clarity in application of the SOCI Act.

[21] Optus submission on exposure draft: Security Legislation Amendment (Critical Infrastructure) Bill 2020 (homeaffairs.gov.au)

Therefore, if a risk-based approach to regulating AI is adopted, the CSCRC supports the production of detailed supporting guidance to underpin implementation and compliance.

16. Is a risk-based approach better suited to some sectors, AI applications or organisations than others based on organisation size, AI maturity and resources?

A risk-based approach is better suited to larger organisations that are investing in the implementation and development of AI technologies and have the resources to support risk management and assessment, which can be costly and time consuming. As noted above, a threshold limit could be applied to regulate compulsory transparency reporting as it relates to a private entity's use of AI technologies.

17. What elements should be in a risk-based approach for addressing potential AI risks? Do you support the elements presented in Attachment C?

The CSCRC supports the elements presented in Attachment C in principle. More detailed information as to the operation and implementation of these elements will be required for further consultation if a risk-based regulatory approach is pursued by government.

18. How can an AI risk-based approach be incorporated into existing assessment frameworks (like privacy) or risk management processes to streamline and reduce potential duplication?

The CSCRC has identified two key areas through which an AI risk-based approach could be incorporated into existing frameworks to avoid duplication. These are in relation to:

- Privacy Act – Privacy Impact Assessments; and
- SOCI Act – Critical Infrastructure Risk Management Program

Privacy Act – Privacy Impact Assessment

Australian Privacy Principle 1 (APP 1) requires APP entities take reasonable steps to implement practices, procedures and systems that will ensure compliance with the APPs and enable them to deal with enquiries or complaints about privacy compliance. A privacy impact assessment (PIA) is a systematic assessment of a project that identifies the impact that the project might have on the privacy of individuals, and sets out recommendations for managing, minimising or eliminating that impact[22]. The CSCRC submits that AI-related issues and the use of AI technologies in a relevant project could be captured as part of PIAs, helping drive increased transparency and stewardship of personal data being used in the context of AI.

[22] Guide to undertaking privacy impact assessments | OAIC

Security of Critical Infrastructure Act – Critical Infrastructure Risk Management Program

The Critical Infrastructure Risk Management Program (CIRMP) is intended to uplift core security practices that relate to the management of certain critical infrastructure assets. It aims to ensure responsible entities take a holistic and proactive approach toward identifying, preventing and mitigating risks. The CSCRC submits that, as part of CIRMP reporting, captured entities could be required to report risks that may arise through the use of AI in critical assets.

20. Should a risk-based approach for responsible AI be a voluntary or self-regulation tool or be mandated through regulation? And should it apply to:

a. public or private organisations or both?

The CSCRC submits that a risk-based approach for responsible AI be mandated through regulation, for both public and private organisations. In terms of public sector regulation, the CSCRC submits that all Federal Government departments and agencies should operate using the same risk-based approach as private sector organisations, with annual transparency reporting a mandatory requirement. In relation to private sector regulation, the CSCRC believes that taking a threshold approach based on annual turnover, as outlined in response to Question 9, would assist in implementing any regulatory approach. For private entities that fall below a mandated threshold, voluntary regulation and self-regulation should be encouraged and fostered via outreach and clear guidance.