Public Consultation on Safe and Responsible AI in Australia

Submission from the ARDC

The ARDC welcomes the opportunity to contribute comments for the public consultation on safe and responsible AI in Australia. Both the Discussion Paper and the Rapid Response Information Report - *Generative AI: Language Models and Multimodal Foundation Models* - provide a helpful introduction to emerging artificial intelligence systems and the challenges, risks and opportunities posed by the application of these technologies. Below, we offer responses to selected consultation questions and put forward our set of recommendations.

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

There is no enforceable AI-specific regulation in Australia. As discussed in the NSTC Report and Discussion Paper, Australian laws can be applied to address some of the risks posed by AI. e.g. privacy laws (data sharing and personal info in training data), consumer protection laws (product liability obligations and claims), copyright and intellectual property laws (contents used by GenAIs), and anti-discrimination laws (bias in decision making and exclusion/denials). While these laws provide protection as described above, there is no single integrated law that addresses AI specifically. Can existing laws be applied or adapted to meet the regulatory challenges posed by AI? Or should the government introduce new overarching regulations for AI in particular? In the process of addressing these questions, it will be necessary to avoid AI exceptionalism ('the idea that AI is riskier or requires greater protection, just because it is AI'[1]). Furthermore, we emphasize the need to examine the unique characteristics of AI as a socio-technical system, the capability of AI as an actor (cognizant or not) and the risk profile that arises due to these features. Other examples of gaps include the lack of clear articulation of responsibilities and the lack of legal mechanisms for an affected individual to access information regarding an algorithm when seeking redress.

That said, the above questions also imply that the potential risks of AI technologies can be anticipated and, to some extent, mitigated by filling in the "gaps" within the existing regulatory landscape.[2] Yet, as the Robodebt Scheme clearly showed, not all the risks of novel technologies can be accurately foreseen and effectively evaluated beforehand. Professor Enrico Coiera and his colleagues have warned that "it is the unintended consequences of these technologies that we are truly unprepared for."[3] In a paper on AI in healthcare, they noted that "there is currently no national framework for an AI-ready workforce" or "overall regulation of safety."[4] If the federal government and stakeholders in the AI sector are committed to supporting "safe and responsible AI," then it will be necessary to take seriously the lessons and learnings from the past.

---

[1] DeCamp, Matthew, and Charlotta Lindvall. "Mitigating bias in AI at the point of care." *Science* 381, no. 6654 (2023): 150-152.

[2] DISR, *Safe and responsible AI in Australia* (Discussion Paper, June 2023) 34.

[3] Coiera, Enrico W., Karin Verspoor, and David P. Hansen, 'We need to chat about artificial intelligence' (2023) *Medical Journal of Australia* 1.

[4] As above.

As commentators have noted, the Robodebt Scheme was not an artificial intelligence system.[5] Nevertheless, it does serve as a cautionary tale for communities in Australia and abroad grappling with the challenges of rapid developments in ADM and AI technologies. AI expert Professor Anton van den Hengel wrote,

> Automation of some administrative social security functions is a very good idea, and inevitable. The problem with Robodebt was the policy, not the technology. The technology did what it was asked very effectively. The problem is that it was asked to do something daft.[6]

In a recent op-ed published in *The Age*, Rachael Falk, CEO of the Cyber Security Cooperative Research Centre, warned that "Another robo-debt disaster's inevitable if we give AI too much autonomy."[7] She argues that "AI is only as good as the algorithm that operates it, the data that trains it and the laws that underpin it."[8] Ms Falk also referenced a number of recommendations from the Final Report of the Royal Commission into the Robodebt Scheme. Key recommendations include those concerning automated decision making. The Royal Commission proposed the "establishment of a body to monitor and audit automated decision-making"[9] and called for legislative reform of the application of ADM in the public sector. Recommendation 17.1 reads as follows:

> The Commonwealth should consider legislative reform to introduce a consistent legal framework in which automation in government services can operate.
>
> Where automated decision-making is implemented:
> - there should be a clear path for those affected by decisions to seek review
> - departmental websites should contain information advising that automated decision-making is used and explaining in plain language how the process works
> - business rules and algorithms should be made available, to enable independent expert scrutiny.[10]

---

[5] Michelle Lazarus and Joel Townsend, 'Automation, uncertainty and the Riobodebt Scheme' *Monash Lens* (23 March 2023) <https://lens.monash.edu/@michelle-lazarus/2023/03/22/1385582/automation-uncertainty-and-the-robodebt-scheme>.

[6] Quoted in Adam Graycar and Adam Master, 'The Robodebt scheme failed tests of lawfulness, impartiality, trust and integrity,' *The Conversation* (3 November 20212) <https://theconversation.com/the-robodebt-scheme-failed-tests-of-lawfulness-impartiality-integrity-and-trust-193832>.

[7] Rachael Falk, 'Another robo-debt disaster's inevitable if we give AI too much autonomy', *The Age* (12 July 2023) <https://www.theage.com.au/technology/another-robo-debt-disaster-s-inevitable-if-we-give-ai-too-much-autonomy-20230712-p5dnlv.html>.

[8] As above.

[9] Recommendation 17.2: Commonwealth of Australia, *Royal Commission into the Robodebt Scheme Report* (Volume 1) (2023) xvi.

[10] As above.

The above proposals recognise the central role of humans in decision-making and judgment and the need for systemic reform and consistent regulation of automation in the Commonwealth public sector. When read in context with other recommendations from the Final Report, Recommendations 17.1 and 17.2 inform efforts to rebuild public trust in government services and public administration.

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

Due to the complexities of AI systems, a multi-faceted, whole-of-society approach is required to supplement the work on AI regulation. Below, we canvas ideas for a national framework for responsible AI followed by comments on data quality, data management and algorithmic transparency; best practices in public funding for AI research; and national research infrastructure.

(i) Towards a national framework to support safe and responsible AI

While regulatory approaches have been proposed to manage AI risk, regulation alone is not sufficient. In contrast, holistic and systemic approaches that may include regulatory framework, consider the broader context of AI, enable proactive risk identification and mitigation, and promote ethical and sustainable AI governance. These approaches to managing AI risk consider the structural issues and wider social, economic and political context in which AI operates.[11] Systemic responsible AI practices encompass a broader set of considerations, including economic risks, job displacement, liability, and reputation risks, and emphasize the importance of responsible AI practices, such as governance, ethically designed solutions, risk control, and training and education, in addressing AI risks.[12] These practices go beyond regulatory compliance and promote a holistic approach to AI risk management.

In this context, Australia has already embarked on voluntary AI-specific governance through the AI Ethics Framework, which would be one of the components of a systemic approach to AI. The concern however would be that a regulatory driven framework would not just stifle innovation,[13] but would also create a regulatory ceiling effect.[14] At the same time, without adequate oversight, AI applications may be misused, resulting in unintended harm. Therefore, it is imperative to design a system that includes voluntary schemes, technical standards, systems approach to risk management, self-regulations, AI-regulations, regulatory collaborations, governance, by-design-considerations, and so on. An important consideration from the research sector would

[11] Schim van der Loeff, Agnes, Iggy Bassi, Sachin Kapila, and Jevgenij Gamper. "AI Ethics for Systemic Issues: A Structural Approach." *arXiv e-prints* (2019): arXiv-1911. https://doi.org/10.48550/arxiv.1911.03216

[12] Wang, Yichuan, Mengran Xiong, and Hossein Olya. "Toward an understanding of responsible artificial intelligence practices." In *Proceedings of the 53rd hawaii international conference on system sciences*, pp. 4962-4971. Hawaii International Conference on System Sciences (HICSS), 2020. https://doi.org/10.24251/hicss.2020.610

[13] Fernandes Da Silva Ranchordas, Sofia Hina. "Experimental Regulations for AI: Sandboxes for Morals and Mores." *Morals & Machines*,1 (2021): 86-100. https://doi.org/10.5771/2747-5182-2021-1-86

[14] Short, Jodi L., and Michael W. Toffel. "Making self-regulation more than merely symbolic: The critical role of the legal environment." *Administrative Science Quarterly* 55, no. 3 (2010): 361-396.

be exempt research activities through a sand-boxing approach[15] and self-governance where appropriate.

(ii) Data Quality, Data Management and Algorithmic Transparency

At the heart of AI risk are data and algorithms. Strengthening data quality management and improving algorithm transparency are crucial for reducing algorithm bias and controlling risks.[16] FAIR is a set of principles that set out the rationale and best practice for making data and other digital objects Findable, Accessible, Interoperable and Re-usable. The adoption of these internationally recognised principles for Australian AI data and algorithms would be a significant contribution to the ethics and trustworthiness of AI.

Formal data quality frameworks also exist for other purposes which could be adopted for the development of AI. For example, the Australian Bureau of Statistics' Data Quality Framework (ABS DQF) is designed "for use by a range of data users and providers in different settings, including government agencies, statistical agencies and independent research agencies."[17] The ABS DQF was agreed to be used "to assess the quality of performance indicator data linked to a number of National Agreements in key policy areas signed by the Council of Australian Governments (COAG) in late 2008."[18] It would be possible to design a framework to assess the quality of data used as input to AI models.

The Australian Research Data Commons has a number of infrastructure programs to support FAIR and quality in the context of AI in research.

(iii) Public Funding for AI Research

We propose that DISR investigate best practices in government and philanthropic funding of AI research. The outcome of this work could inform the design of future funding programs and assist coordination activities among Australian funders and their stakeholders. The Discussion Paper briefly discussed recent Commonwealth AI funding initiatives. Consistent with the revised List of Critical Technologies in the National Interest, the federal government has allocated funding in 2023-24 Budget to support the work of the National Artificial Intelligence Centre and the implementation of the Responsible AI Adopt Program. This funding builds upon investments in AI technologies made by the prior government. We note that funding support for AI research has also been provided by the Medical Research Future Fund (MRFF). As part of its National Critical Research Infrastructure Initiative, the MRFF awarded five grants for "applied artificial intelligence research in health."[19] The successful grants were announced by the former Minister

---

[15] Pošćić, Ana, and Adrijana Martinović. "Regulatory Sandboxes under the Draft EU Artificial Intelligence Act: An Opportunity for SMEs?." *InterEULawEast: Journal for the International and European law, economics and market integrations* 9, no. 2 (2022): 71-117. https://doi.org/10.22598/iele.2022.9.2.3

[16] Zhang, Jie, and Zong-ming Zhang. "Ethics and governance of trustworthy medical artificial intelligence." *BMC Medical Informatics and Decision Making* 23, no. 1 (2023): 7 https://doi.org/10.1186/s12911-023-02103-9.

[17] ABS, 1520.0 - ABS Data Quality Framework, May 2009 <https://www.abs.gov.au/ausstats/abs@.nsf/mf/1520.0>.

[18] As above.

[19] Australian Government, *Funding for research projects using Artificial Intelligence technologies and methodologies* (2019) <https://business.gov.au/grants-and-programs/applied-artificial-intelligence-research-in-health>.

of Health, Greg Hunt, in June 2020.[20] In addition, AI and machine learning were both discussed in the *2021 National Infrastructure Roadmap*. We anticipate that both will also feature in the forthcoming National Digital Research Infrastructure Strategy.

Government funding for AI research in Australia is modest when compared with investments made by multinational corporations and other developed countries. In June 2023, the UK Secretary of State for Science, Innovation and Technology, Chloe Smith, announced an investment of £54 million for AI research and the development of a data science workforce.[21] Secretary Smith's announcement was coupled with the launch of the UK Geospatial Strategy 2030. The strategy underpinning the government's funding announcement is to drive innovation and economic growth by harnessing the potential of AI, satellite imaging and real-time data. The UK approach illustrates the importance of supporting AI research in relation to other government priorities (e.g. decarbonisation and the transition to renewable energy) and investment in related technologies (e.g. location data and technologies). The UK announcement on AI investment was preceded by a funding allocation of £100 million to establish the Foundational Model Taskforce in April. The aim of the expert taskforce is to ensure sovereign capabilities and the broad adoption of safe and reliable foundation models in the UK. Given developments abroad, Australian funding bodies face a number of challenges. How can the return from local public investment in AI technologies be maximised? How can AI grant programs be designed in a way that can complement research in the private sector? Which policies are generally applicable to all types of AI research and which are sector-specific? Coordination among Australian government and philanthropic funding organisations could minimise duplication and facilitate transdisciplinary research necessary to address complex societal challenges.

(iv) National Research Infrastructure

And yet, discrete research projects alone would not be sufficient to create enduring capability for addressing the risks of AI. Complementing research projects with research infrastructure offers a pathway to increase the return on investment of research, avoid duplication, and maintain ongoing reliable and underpinning capabilities.

Developments in research infrastructure offer critical and innovative tools to conduct and support such research, from conception to publication. In particular, digital research infrastructure (DRI) can be designed to facilitate the important task of collecting, processing, storing and providing access to cultural and scientific data "in accordance with the principles of good data governance and stewardship."[22] While AI research activities tend to be undertaken as discrete projects, research infrastructure provides longer term capability for STI (science,

[20] The list of successful grants can be found in the former Health minister's media release. Department of Health and Aged Care, '$19 million for Artificial Intelligence health research projects,' (Media release), 29 June 2020 <https://www.health.gov.au/ministers/the-hon-greg-hunt-mp/media/19-million-for-artificial-intelligence-health-research-projects>.

[21] UK Department of Science, Innovation and Technology, *£54 million boost to develop secure and trustworthy AI research*, (Press release), 14 June 2023 <https://www.gov.uk/government/news/54-million-boost-to-develop-secure-and-trustworthy-ai-research>.
https://www.gov.uk/government/news/54-million-boost-to-develop-secure-and-trustworthy-ai-research

[22] UNESCO, *Recommendation on Open Science* (2021) cl 7(b), 10.

technology and innovation) to address complex society challenges in an ongoing and coordinated manner. As the *Brno Declaration* observed, research infrastructure (RI) "brings together a critical mass of human, material and financial resources, and integrates a large variety of stakeholders … in a multi-disciplinary and cross-sectorial way."[23] Furthermore, RIs

> … are essential to conduct breakthrough fundamental research and drive excellence, extending the frontiers of human knowledge beyond yet known horizons, to perform cutting-edge applied research, advancing the technology development and feeding into science-driven innovation, and to provide services in the support of tackling grand societal challenges…[24]

The National Collaborative Research Infrastructure Strategy (NCRIS) is an existing Australian Government initiative to deliver such national research infrastructure capabilities. Among the 24 existing facilities several provide system wide support for data and compute (eg the Australian Research Data Commons, the National Computational Infrastructure, and the Pawsey Supercomputing) and many have domain or problem specific analytics capabilities (eg the Atlas of Living Australia, the Integrated Marine Observation System). ARDC has pointed out some examples in this submission where its national infrastructure capability helps address the issues of data quality and integrity and algorithm transparency.

Research infrastructures are a key component of a national plan to support responsible AI practices in Australia and the Australian Goverment's NCRIS strategy should be leveraged.

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 development and uptake of AI in Australia.

In managing AI, it is important to have a collaborative approach to AI risk management. This requires coordinating resources, expertise and energy to address the challenges. Given the high stakes that may be involved in AI-driven decision-making, a multi-stakeholder AI governance framework to coordinate and facilitate between regulators, industry stakeholders, researchers, and civil society organisations, is essential.[25] By incorporating diverse perspectives and expertise, regulatory agencies can develop comprehensive and effective strategies for AI risk management.

Regulatory agencies should collaboratively establish guidelines and regulations that ensure the quality and integrity of data used in AI systems. This can also improve trust between players, improve accountability, transparency and traceability and can further enhance and enable effective risk management.

---

[23] The full title of this source is the *Brno Declaration on Fostering a Global Ecosystem of Research Infrastructure*: European Strategy Forum on Research Infrastructures (ESFRI), *Brno Declaration on Fostering a Global Ecosystem of Research Infrastructure* (14 December 2022) clause 2, 2 <https://www.esfri.eu/latest-esfri-news/brno-declaration-ris#:~:text=The%20Brno%20Declaration%20on%20Fostering,the%20closing%20of%20ICRI%202022.>.

[24] As above, clause 1, 2.

[25] Aloisi, Antonio, and Valerio De Stefano. "Between risk mitigation and labour rights enforcement: Assessing the transatlantic race to govern AI-driven decision-making through a comparative lens." *European Labour Law Journal* 14, no. 2 (2023): 283-307. https://doi.org/10.1177/20319525231167982

To some extent, the NCRIS facilities provide coordination across research and within different scientific domains on the application of AI. We would welcome a nationally coordinated initiative across the whole economy in which to place our work.

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

Given the international context of science and the digital economy, the coordination of AI governance in Australia will need to take account of, and be responsive to, emerging global norms and standards. Ideally, local governance should be designed in a way that it can easily integrate into global regulatory systems and policy forums. In this regard, the adoption of the ISO definition of AI provides a helpful starting point. In a recent co-authored paper with the National AI Centre, Standards Australia have documented other key ISO publications and ongoing standards projects.[26]

The Discussion Paper makes reference to the regulatory sandbox administered by ASIC in the context of financial services. According to Pošćić & Martinović (2023), regulatory sandboxes have emerged as an alternative avenue for risk management of disruptive technologies, including AI.[27] By creating a controlled environment for testing and experimentation, regulatory sandboxes enable research and development thereby fostering innovation under real-life conditions. Such a tool also allows regulators to understand the risks and challenges associated with AI technology. We recommend that DISR examine the regulatory sandboxes under the draft EU AI Act. To facilitate experimentation and innovation, we support the introduction of AI regulatory sandboxes informed by overseas experiences and building upon the work of ASIC in financial services.

Similarly to many other countries, Australia is a signatory to the *UNESCO Recommendation on Ethics in Artificial Intelligence* (2021). UNESCO illustrated the application of their recommendation in a recent publication titled, *Foundation Models such as ChatGTP through the Prism of the UNESCO Recommendation on Ethics in Artificial Intelligence* (2023).[28] In this paper, UNESCO reports that they are partnering with Microsoft and Telefonica to implement its Recommendation.[29] They also indicated that they plan to launch the Observatory on the Ethics of AI with the Alan Turing Institute. As part of its work on AI, UNESCO hosted the first annual Global Forum on the Ethics of AI in December 2022. This forum is designed for sharing global best practices among AI stakeholders worldwide. Thus, as the above discussion illustrates, UNESCO has been particularly active in the area of AI. The recent return of the US to UNESCO also

[26] Standards Australia and CSIRO National Artificial Intelligence Centre, *Introduction to Standards for Artificial Intelligence* (May 2023) 9, available at <https://www.standards.org.au/news/unlocking-the-potential-of-ai-how-responsible-ai-can-drive-innovation>.

[27] Pošćić, Ana, and Adrijana Martinović. "Regulatory Sandboxes under the Draft EU Artificial Intelligence Act: An Opportunity for SMEs?." *InterEULawEast: Journal for the International and European law, economics and market integrations* 9, no. 2 (2022): 71-117. https://doi.org/10.22598/iele.2022.9.2.3

[28] UNESCO, *Foundation Models such as ChatGTP through the Prism of the UNESCO Recommendation on Ethics in Artificial Intelligence* (June 2023) <https://unesdoc.unesco.org/ark:/48223/pf0000385629>.

[29] As above, 3.

highlights the significance of UNESCO as a forum for the development of global standards on AI and other technologies.

In addition, we note recent announcements made by prominent figures from the UK and US. In July 2023, the UK Foreign Secretary Jame Cleverly called for the global governance of AI at the first United Nations Security Council session on artificial intelligence.[30] Mr Cleverly emphasised the need for global cooperation to ensure that AI technologies and the rules governing their application are developed in a way which benefits communities around the world. The UK is also due to host a global summit on AI safety in autumn. Further, US President Joe Biden recently announced that seven tech companies, including Google, Amazon, Meta and Microsoft, agreed to meet a set of AI safeguards set by the Whitehouse.[31]

Conclusion

As noted above, there is no enforceable AI-specific regulation in Australia. Existing Australian laws can be applied or adapted to address some of the risks posed by AI. The lack of AI specific legislation may result in some gaps due to the unique characteristics of AI. That said, regulation is not the only approach to managing AI. We propose that DISR explore alternative approaches which can incorporate a repertoire of systems, tools and techniques.

While new technologies may enable automation in decision-making, it does not necessarily follow that we *should* adopt such technologies. A key question will be what is the role of humans given the possibility of automated or augmented decision making? In this submission, we offered comments in response to four consultation questions drawing upon our work in the area of digital research infrastructure. We proposed that the recommendations on automated decision-making from the Final Report of the Royal Commission into the Robodebt Scheme (Rec 17.1 and 17.2) serve as one of the key references for the development of AI governance in Australia. A critical focus on humans and human agency are necessary to establish systems of accountability and to work towards the development of human-centred artificial intelligence systems. In addition, we also discussed some of UNESCO's work on AI, the UK's call for global governance of AI and suggested an investigation into best practices in government and philanthropic funding for AI research.

[30] UK Government, *Foreign Secretary to call for international cooperation to manage global implications of Artificial Intelligence*, (Press release, 18 July 2023) <https://www.gov.uk/government/news/foreign-secretary-to-call-for-international-cooperation-to-manage-the-global-implications-of-artificial-intelligence>.

[31] Matt O'Brien and Zeke Miller, *Amazon, Google. Meta, Microsoft and other tech firms agree to AI safeguards set by the Whitehouse*, AP News, 21 July 2023, <https://apnews.com/article/artificial-intelligence-safeguards-joe-biden-kamala-harris-4caf02b94275429f764b06840897436c>.

Recommendations

(i) That the recommendations on automated decision-making from the Final Report of the Royal Commission into the Robodebt Scheme (Rec 17.1 and 17.2) serve as one of the key references for the development of AI governance in Australia.

(ii) We recommend that DISR examine the regulatory sandboxes under the draft EU AI Act.

(iii) We propose that DISR investigate best practices in government and philanthropic funding for AI research to inform the design of future funding programs in Australia.

(iv) To complement AI regulatory reforms, DISR should consider support for data quality, data management, and algorithmic transparency, potentially in collaboration with NCRIS research infrastructure.

(v) We recommend that DISR adopts a collaborative and coordinated approach to AI risk management, and consider how existing NCRIS facilities and services could be leveraged for the purposes of a national plan to support responsible AI practices in Australia.

About the ARDC

The ARDC drives the development of national digital research infrastructure that provides Australian researchers with a competitive advantage through data. The ARDC is Australia's peak body for research data. We aim to accelerate research and innovation by driving excellence in the creation, analysis and retention of high-quality data assets. We facilitate access to national digital research infrastructure structure, platforms, skills, data sets and tools from academia, industry and government for all Australian researchers. The ARDC is funded through the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS) to support national digital research infrastructure for Australian researchers.

The contact for this submission is Dr Adrian Burton, Deputy CEO, Director, People RDC. He can be contacted via email or phone: <adrian.burton@ardc.edu.au>; (02) 6125 6659.