Safe and Responsible AI in Australia

RMIT input to Australian Government Discussion Paper

July 2023

RMIT University welcomes the opportunity to comment on the Safe and Responsible AI in Australia Discussion Paper. Our feedback includes the views of a range of key stakeholders, including research leaders, researchers and research management professionals realising RMIT's mission, goals and strategy for achieving research and innovation with positive impact.

Central to this is the ethical and responsible design, development, deployment and use of enabling technologies, including artificial intelligence (AI). Recognising both the opportunities and challenges that AI technologies present, as well as the rapid advancement and developments of these technologies.

Also recognising the unique position of universities and research communities, as being users of these technologies, as well as actively involved in their design, development, deployment and evaluation, in conducting research to "… generate new concepts, methodologies, inventions and understandings" (Australian Code for the Responsible Conduct of Research 2018).

Further details are provided below, ordered in keeping with the subheadings and grouped questions presented in the Discussion Paper.

Definitions

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

Yes. Noting these definitions of key terms are drawn from International Organisation for Standardization (ISO) definitions and accord with the understanding and use of these key terms in a higher education setting. It was suggested that the description of a large language model (LLM) as a "type of generative AI that specialises in the generation of human-like text" be corrected to "a probabilistic model that aims to represent the patterns of language usage in human text" for accuracy and in keeping with the understanding and use of this term in computing sciences.

Potential gaps in approaches

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?**

Recognising Australia's world-leading research and innovation capabilities, RMIT stakeholders identified potential risks relating to AI and research integrity. These risks include research falsification, proliferation of misinformation, decreased transparency, lack of reproducibility, entrenchment of bias, expansion of the utility of AI and risks relating to breaches or sharing of confidential, sensitive, or private data in research.

RMIT stakeholders did not have any specific suggestions for possible regulatory action to mitigate these research integrity related risks. However, RMIT stakeholders observed that the Australian Code for the Responsible Conduct of Research 2018 is well-established and being a principles-based framework facilitates adaptation and application across the breadth and depth of research disciplines, as well as to novel challenges in a research context, more comfortably and quickly than a rules-based approach.

RMIT stakeholders also identified potential gaps in relation to the ability or rights of people to appeal decisions made or supported through Automated Decision Making (ADM), including the ability or right to be excluded from a process administered by ADM and the ability or right to appeal decisions made by ADM, and these rights being enshrined in legislation.

RMIT stakeholders also observed in an education context that AI technologies presents particular risks to the teaching and learning of knowledge skills, as well as over-reliance on generative AI to outsource and automate knowledge work. These risks are not currently acknowledged or outlined in the Paper.

RMIT stakeholders also identified potential risks from AI relating specifically to recognition of and respect for Indigenous knowledges, knowledge systems and cultures. In an Australian context, this assumes particular importance with respect to Aboriginal and Torres Strait Islander knowledges, knowledge systems and cultures. Suggested strategies to mitigate these risks included deep and meaningful engagement and co-design of AI with Indigenous peoples and communities to understand and their perspectives and ensure respect for Indigenous knowledges, knowledge systems and cultures in the ethical and responsible design, development, deployment and use of AI, as well as to further encourage the implementation of Indigenous knowledge and knowledge systems in the curriculum.

This is also a consideration for culturally and linguistically diverse (CALD) people and communities in Australia, with various AI technologies displaying inherent cultural biases.

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.**

RMIT stakeholders were supportive of non-regulatory initiatives which would enhance and expand communities of practice and networks across the higher education, industry, government and community sectors, with a view to establishing, sharing and monitoring ethical and responsible AI practices, as well as broader public awareness and education, with a view to increasing public trust.

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. Responses suitable for Australia**

The current regulatory landscape appears disparate and disjointed, with numerous general regulations and specific regulations addressing aspects of AI technologies governance and numerous regulatory agencies, bodies and departments at Commonwealth and State level having limited and in some cases duplicating and/or overlapping accountabilities.

The establishment of an overarching regulatory framework which is risk-based and principles-based and which considers and sits over and across, or replaces, existing general regulations or sector-specific laws, along with the establishment of a central and suitably expert regulatory body, could assist in coordination of AI governance across government.

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?**

RMIT stakeholders did not suggest any other governance measures being taken or considered by other countries which are not discussed in this paper, as being relevant, adaptable and desirable for Australia. RMIT stakeholders indicated a preference for governance measures which were proportionate to both the risks and benefits presented by AI, which adopt a principles-based approach and in which the wellbeing of humans and society, animals and the environment are paramount.

RMIT stakeholders observed that there are many similarities between gene technologies and AI technologies, with both these technologies presenting significant challenges and opportunities and both having undergone periods of rapid advancement. It was suggested that Australia's regulatory approach to gene technologies, including the establishment of the gene technology framework and a central Commonwealth regulator, could be relevant, adaptable and desirable in relation to AI technologies and in an Australian context.

Target areas

6. **Should different approaches apply to public and private sector use of AI technologies? If so, how should the approaches differ?**

As above, RMIT stakeholders were of the view that the regulatory approach to AI in Australia should be proportionate to both the risks and benefits presented by AI, adopt a principles-based approach and apply consistently to the design, development, deployment and use of AI technologies. This proportionate and risk-based approach would ideally be principles-based and lend itself to being applied broadly across all sectors.

However, it may be appropriate to develop regulations specific to particular AI technologies, and/or particular activities or dealings with these technologies and/or to particular sectors. In this regard, an RMIT stakeholders observed the lack of recognition of the role of universities and other research institutions, with one commenting:

"To draw an analogy with the medical space, research is essential for progress, but rigorous evaluation and safety protocols are required before research findings are deployed into the public sphere. Similarly, to ensure that Australia can maintain a leading position in terms of the advancement of AI and related technologies, research institutions should not be constrained from pursuing rigorous investigations. However, to ensure the beneficence and safety of such technologies, it be reasonable to require rigorous assessments and requirements before they are released for wide-ranging public use, outside research environments."

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

Suggestions received included:

- Strengthening governance frameworks and practices for the responsible and safe design, development, deployment, and use of AI technologies,
- Providing ongoing training and education that promotes and supports responsible and safe design, development, deployment, and use of AI technologies, and
- Ensuring public servants have the appropriate skills, qualifications, and resources to responsibly

and safely design, develop, deploy and use AI technologies.

8. **In what circumstances are generic solutions to the risks of AI most valuable? And in what circumstances are technology-specific solutions better? Please provide some examples.**

No feedback.

9. **Given the importance of transparency across the AI lifecycle, please share your thoughts on:**
   a) **where and when transparency will be most critical and valuable to mitigate potential AI risks and to improve public trust and confidence in AI?**
   b) **mandating transparency requirements across the private and public sectors, including how these requirements could be implemented.**

RMIT stakeholders agree that transparency is critically important across the AI lifecycle. In the context of mitigating potential risks and improving public trust and confidence, transparency is particularly valuable in the design and development stages, as well as in reporting on these technologies in a way that is open, responsible, and accurate, and also in identifying, disclosing, and managing any related conflicts of interest. For example, the majority of Large Language Models (LLMs) are opaque and do not allow for meaningful human understanding and inspection of how these AI technologies work. Encouraging or requiring the sharing and training sets and model reporting (with the use of model cards for model reporting a good and relevant example https://dl.acm.org/doi/pdf/10.1145/3287560.3287596), assessment and evaluation results, and communicating explanations, inclusive of knowns and unknowns, would provide greater transparency and improve trust.

10. **Do you have suggestions for:**
    a) **Whether any high-risk AI applications or technologies should be banned completely?**
    b) **Criteria or requirements to identify AI applications or technologies that should be banned, and in which contexts?**

No suggestions on specific high-risk AI applications or technologies that should be banned completely, noting those technologies already identified as unacceptable risks. However, RMIT stakeholders recognised the need for rigorous oversight and regulation of high-risk AI applications or technologies such as medical or military, and potentially to pause or ban these where they present real and significant risks to the well-being of humans and society, animals and the environment.

11. **What initiatives or government action can increase public trust in AI deployment to encourage more people to use AI?**

Please see feedback for question 3 above.

Implications and infrastructure

12. **How would banning high-risk activities (like social scoring or facial recognition technology in certain circumstances) impact Australia's tech sector and our trade and exports with other countries?**
13. **What changes (if any) to Australian conformity infrastructure might be required to support assurance processes to mitigate against potential AI risks?**

No feedback.

Risk-based approaches

14. **Do you support a risk-based approach for addressing potential AI risks? If not, is there a better approach?**
15. **What do you see as the main benefits or limitations of a risk-based approach? How can any limitations be overcome?**
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?**
17. **What elements should be in a risk-based approach for addressing potential AI risks? Do you support the elements presented in Attachment C?**
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?**
19. **How might a risk-based approach apply to general purpose AI systems, such as large language models (LLMs) or multimodal foundation models (MFMs)?**
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?**
    b) **developers or deployers or both?**

As discussed above, RMIT is supportive of a proportionate and risk-based approach to both address the potential risks and realise potential opportunities presented by AI technologies, and in which the wellbeing of humans and society, animals and the environment are paramount. That is, the amount and type of regulation would increase proportionate to the risks presented, and with the wellbeing of humans and society, animals and the environment being paramount—meaning some AI technologies or uses of AI technologies may be banned or paused where they present unacceptable risks.

No feedback on questions 15 or 16.

RMIT stakeholders broadly supported the elements for the risk-based approach presented in Attachment C. Noting, that the impact assessment would also benefit from consideration of potential benefits in addition to potential risks and a justification for how any risks identified are justified. Stakeholders also observed that the concept of “human in the loop” in this context (that is, seemingly to represent a notion of human oversight and/or assessment) was essential but that the phrasing was problematic. In that “human in the loop” is often used to describe scenarios where humans are working or collaborating with an AI system to achieve a goal. That is, they are part of the process, which is substantially different from notions of human oversight and assessment as it is being used in this context. For clarity, it was suggested using terms such as “human oversight” or “human assessment” in this context, rather than “human in the loop.”

No feedback on questions 18 or 19.

In closing, RMIT stakeholders see a need for and value in Australia adopting a common proportionate risk-based and principles-based regulatory approach to promote and foster the safe and responsible design, development, deployment and use of AI in Australia—as well as in these regulations being mandated and applying across all sectors (given the potential risks and recognising self-regulation is not appropriate or justifiable). Also, recognising that there may be a need for more specific and limited regulations to apply to particular AI technologies and/or to particular sectors and/or in particular contexts.

Given the complexity of the current landscape and need to ensure that the optimisation goals of AI-based technologies and systems are congruent with objectives that minimise risks of harm and promote societal and ethical goals and objectives, RMIT stakeholders also emphasised the need for further resourcing for this area, as well as the establishment of a central and suitably expert regulatory body, in addition to the overarching regulatory framework outlined above.

Thank you once again for the opportunity to contribute to this consultation, and we welcome requests for further detail or discussion.

Yours sincerely,

Professor Calum Drummond AO
DSc PhD FNAI(USA) FTSE FAICD FRACI FRSC(UK) CChem
Deputy Vice-Chancellor and Vice-President
Research and Innovation