MONASH University

SAFE AND RESPONSIBLE AI IN AUSTRALIA

August 2023

INTRODUCTION

The rise of Artificial Intelligence is a disruptive technological development that touches on all aspects of life including social, work, and education.

Monash University academics are developing, researching and using AI across a range of disciplines including IT, engineering, law, ethics, industry development and global competition; in our educational practices and governance; and as educational and research leaders in terms of our own organisation response. Our submission to this consultation is drawn from all these perspectives.

We note the following people with particular expertise who are available for further consultation:

- Professor Chris Marsden: Professor of AI and the Law, Associate Director Data Futures Institute
- Professor Geoff Webb: Professor of Data Science and AI
- Associate Professor Tim Fawns: Monash Education Academy
- Dr Paul McIntosh: Innovation Lead Strategic Enablement and Adjunct Senior Research Fellow Opportunity Tech Lab, Monash Business School

We also recommend the Association for Computing Machinery's FAccT Network which stores the papers and presentations of numerous conferences and workshops on matters pertinent to fairness, accountability and transparency in the development and use of AI tools including socially responsible algorithmic decision making, software fairness and trust and reliance in AI assisted tasks. The Association for Computing Machinery Technology Policy Council has also released a new guideline on GenAI.

The recent increased availability of generative AI (GenAI) prompted action to prepare the Monash community for the opportunities and challenges GenAI tools present to teaching, learning, assessment, research methods and academic integrity across all courses include research degrees. The work undertaken by the Monash GenAI Task Force informed our response to the AI in Education Inquiry by the House of Representatives Education, Employment and Training standing committee.

The discussion paper, *Safe and Responsible AI in Australia*, addresses most of the issues and concerns raised about the responsible use of AI and we support the general approach to regulation and consumer protection.

We note the focus of this consultation is to identify potential gaps in the existing governance environment along with any extra mechanisms required to support the development and adoption of AI.

It is Monash's view that:

- it is right that government is reflecting on risks and responsible use
- responsible use applies to the creator/supplier, the value chain of often global developers, and to the user
- AI is a means rather than an end, with uses in all aspects of our lives, and the likelihood of becoming ubiquitous in many professions, industries and social settings. The scale of the technology and global application make it too big for the Australian government to regulate in total, as with the Internet.

This leads us to conclude that regulation should be focused on a risk-based assessment of the application and context rather than the existence of AI tools, in accord with the European Union and Council of Europe, and the G7 Hiroshima AI Process.

Threats and opportunities are real. Managing the risks and reaping the benefits will require a considerable investment of resources to achieve critical size in human intelligence to analyse and develop AI, inform ethical uses, predict the harms and benefits, and support industrial transitions and productivity improvements in a time of population ageing.

Universities have a critical role to play. It will require a considerable investment of resources to develop the human capital and to train the next generation of Australian educators and researchers. The challenge may be as great as that required to train Internet researchers in the 1990s and Australia cannot afford to become a late adopter.
(The UK government has committed £100 million to an expert taskforce to help the UK build and adopt the next generation of safe AI.)

We respond to the specific questions below and are happy to engage further as required.

QUESTIONS

DEFINITIONS

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

   Some of the definitions would benefit from being more precisely defined, for example:

   - Large Language Model is a type of generative AI trained on a large number of textual sources to generate human-like text in response to a prompt.
   - Multimodal Foundation Model is a type of generative AI trained on a large number of resources such as text, images, numbers and code, that can be easily customised or fine-tuned to different tasks.

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?

   - We note that while algorithmic risk in building LLMs is addressed in the discussion paper, there is a risk of using existing APIs and models such as ChatGPT that are particularly aligned with human preferences and cultural/social norms in North America. These norms are not global and potentially require refinement for the Australian context.
   - The discussion paper does not cover risks associated with the integration of AI into existing work practices and the responsibility of employers. The Work Health and Safety Regulations (WH&S) may cater for this through psychosocial hazards at work[1], however we would like to see it explicitly included. AI and the people using these tools form a sociotechnical system. There is a workplace risk in introducing technology without considering both. Monash has developed guidelines to address emerging technology psychological hazards, as well as physical, in deploying new Virtual Reality technology.
   - The relative constraints that the Rule of Law places on the ways that power may be exercised does not clearly fit within the regulatory framework as proposed. The Rule of Law, and Australia's continued suggestion that it is a state that adheres to that concept, operates as a fundamental limitation to the way that public power can be exercised. The Rule of Law means that power should be exercised in a way that is, amongst other things, predictable and procedurally clear. As is illustrated through this document, there are problems with transparency in the black box. This means that there are real implications for the ways in which AI can be used to exercise power, particularly public power, that go beyond questions regarding Australia's regulatory approaches.

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.

   Governments need to stay abreast of fast-changing AI technology for effective regulation and oversight. However, the lack of current technical expertise and fast pace of change, as evidenced by the Australian chief scientist's hesitance to forecast beyond two years, is concerning. Addressing the skills gap is essential to leverage AI's

[1] For examples of psychosocial hazards at work, refer to https://www.comcare.gov.au/safe-healthy-work/prevent-harm/changes-to-whs-laws

economic potential, tackle ethical issues, and address challenges in transitions to a future where AI is powerful and ubiquitous.

The Australian Government could launch an initiative to build AI expertise, particularly in AI safety, governance and ethics, into government. Initially, these experts could focus on the safety and security of AI deployments in critical sectors like law enforcement, finance, defence, and critical infrastructure.

The most effective way to identify and recruit AI safety experts is through partnerships with Australian universities. AI safety and ethics training programs, combined with policy training (for example through the Monash University Data Futures Institute and associates), could serve as a training and recruitment pathway. We note the Cranlana Centre for Ethical Leadership is developing programs and tools to support leaders regarding the ethical issues raised by AI.

Benefits of this approach include:

- improved decision-making: AI experts can provide informed recommendations that could shape public decision-making with and about AI
- future-readiness: As AI becomes increasingly prevalent across sectors, possessing in-house know-how allows the government and public services to remain on top of developments, control potential hazards, and fully leverage AI opportunities
- economic advantages: A better understanding of AI as a technology and its economic and social impacts can stimulate economic growth by encouraging safe AI adoption across sectors such as industry, healthcare, and education, while mitigating harms or risks from unsafe use.

Additions to the Commonwealth Procurement Rules could prioritise open, interoperable systems that are transparent and non-discriminatory, along the lines of the draft interoperability procurement guidelines put out for discussion by the Digital Health agency.

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.

As with other multi-sector initiatives and innovations, while we note that the creation of another agency or position is not always the answer to coordination, we do recommend the appointment of an AI Commissioner or national coordinator, with specific interest groups/sectors across government, potentially organised by service, for example:

- Business and industry
- Infrastructure
- Health and community services
- Crime prevention, detection and justice
- Science and Technology

We commend the whole of government coordination of the response to foreign interference in higher education. The University Foreign Interference Taskforce (UFIT) successfully brought together university peak bodies and relevant Australian government agencies including Home Affairs, DFAT and Education. The UFIT steering group and sub-groups developed guidelines to counter foreign interference in higher education, in a way that was mostly proportionate to risk and sector relevant. A solution for AI should be nimble enough to respond quickly as the AI environment is rapidly evolving.

To encourage development and take up, we recommend harmonising investment decisions in relation to the funding of infrastructure, research, commercialisation and effect of particular AI systems, tools and uses.

RESPONSES SUITABLE FOR AUSTRALIA

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?

Australia should pay particular regard to the Council of Europe (CoE) Framework Convention on Artificial Intelligence, currently negotiated between Council of Europe members and observers including the United States. This Convention has the potential to become the default global democratic standard for generic AI regulation, as CoE Convention 108+ has for data protection and the Budapest Cybercrime Convention 2001 has for Internet regulation and enforcement cooperation.

These are potentially more important Conventions than the anticipated but not yet existent global AI Treaty that some have called for, and the much delayed European Union draft Regulation on AI (incorrectly known as the 'AI Act'), which we note will not become law until at earliest 2024, and not be enforced until at least 2027[2].

The 'AI Act' itself is not an appropriate instrument for Australia to apply because it is nested in the web of European digital services laws known as the 'digital acquis'. It does contain useful approaches to risk regulation, notably on live facial recognition and the regulation of AI value chains and foundation models, which will be adaptable to the Australian experience.

The United States has continued to pursue sectoral regulation, most notably through the Federal Trade Commission (FTC), as well as a 'soft law' self-regulatory "pact" signed at the White House on July 21. Monash researchers are actively engaged with the National AI Initiative in the White House on these issues[3].

TARGET AREAS

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

   In practice self-regulation doesn't work in profit-motivated fields, but equally, we have seen that public sector uses of technologies can be prone to misuse as well, even if it is inadvertent.

   It might be appropriate for greater latitude in basic research and discovery in AI with greater caution in application through regulation and limits. Regardless of 'sector' the human tendency is to automation bias over human judgement and it is evident that general purpose technologies/platforms can be harnessed for both benefit and harm, e.g. social media. There is a wealth of experience to learn from Big IT and social media and the focus should be on protecting the vulnerable.

   Certain public sectors, including Higher Education, should be encouraged to explore the possibility of building their own large and medium scale language models so as to become less dependent on corporate models and to help address misinformation and some of the normative, discriminatory aspects of public datasets. While there may be concerns about cost, it is possible that the bigger barrier will be the expertise required.

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

   Given the speed of evolution in AI capabilities it will be important to prioritise the development of AI literacy to understand what the pitfalls and consequences might be, especially as relates to bias, privacy, transparency of AI systems. This extends to understanding the limitations of AI, particularly unsupervised tools, and staying alert to the need for human characteristics such as emotion and empathy.

   A paper on public law in a technological era considered the fitness of public law in ensuring and addressing individual rights in the case of government use of AI in decision making[4]. It looked at the regulation of data processing, the transparency of the automated decision making systems and the remedies. It recommended both bottom up and top down responses including the establishment of a whole of government guidance framework, specific privacy reforms in relation to automated decision making (with reference the 2016 General Data Protection Regulation and the CoE Convention 108+) and a range of transparency reforms to the ADJR Act, the AAT Act and the FOI Act.

[2] Monash Professor Chris Marsden's Canberra Times op-ed of 14/I 6/2023 explained some of these steps.

[3] Professor Chris Marsden and the Digital Law Group is continuing intensive research into these proposed laws alongside colleagues from Warwick University (UK), VUB (Free University of Brussels), European University Institute, Centre for European Policy Studies (CEPS), University College London and others. The Digital Law Group can provide specific advice on these issues to the federal government, as they have with DFAT in Geneva at the recent UN AI Summit. They also held expert workshops in Warwick, Brussels and at our Prato Campus in June 2023 to explore any potential policy transfer benefits for Australia and can present our conclusions at any appropriate future forum.

[4] https://www.unswlawjournal.unsw.edu.au/article/revitalising-public-law-in-a-technological-era-rights-transparency-and-administrative-justice, especially Section V Recommendations, Reform and Conclusions.

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.

We know that AI is used for public and private benefit and that the two will not always align. We support government initiatives to ensure all public funded AI tools and uses are disclosed and transparent, and accord with Australia's AI Ethics principles and best practice frameworks. AI tools used in government are public records and should be treated accordingly.

Transparency is important as a foundation of trust and mandates should be required, especially in sensitive situations where decisions are being made about people and their rights to access government services and public-facing services, including "finance, insurance, healthcare & medical research, social care, policing and security, education, transport (AI-guided airliners & automated vehicles), social media, telecommunications.[5]"

Australia should add to global calls for transparency and metadata standards and invest financially to support the development of open access and responsible tools, including those that are more energy efficient. As when universities established Australia's Academic And Research Network (AARNet) in 1989, universities can play a significant role in developing tools tailored for particular contexts, contributing to standards, as well as contributing to standard prompts, use cases etc.

Mandates are difficult in cases of shared production or distributed responsibility, such as AI technologies that involve algorithmic supply chains, and particularly under the technology-as-a-service model. This issue is discussed in the paper Understanding Accountability in Algorithmic Supply Chains[6].

10. Do you have suggestions for:
    a) Whether any high-risk AI applications or technologies should be banned completely?

We support the prohibition of live biometric surveillance on the general public, which has been banned in the European Community via the 'AI Act'. This is a full ban on real-time and retrospective biometric surveillance, emotion recognition and predictive policing systems.

It also expanded the list of prohibited AI practices such as biometric categorization systems using sensitive characteristics (e.g., gender, race, ethnicity, citizenship, religion, political orientation) and untargeted scraping of facial images from the internet or CCTV footage to create facial recognition databases.

    b) Criteria or requirements to identify AI applications or technologies that should be banned, and in which contexts?

We support the EU approach to limit / prohibit systems that exploit people's vulnerabilities, or manipulate human behaviour, or other discriminate access to human rights and public benefits. Consultation across the interest groups as suggested under Q4 and with legal and regulatory experts such as those at Monash will be essential to stay current.

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

Acting instead of reacting is important to promote trust. Rather than waiting for market failure it is critical that governments at all levels engage with industry and interest groups as technologies and uses are developed, with community campaigns that highlight not just what AI can do but how it is already used, and where the benefits are, as well as the limitations and dangers. This could extend to grassroots grants to enable local education and adoption of AI tools for the benefit of particular communities or interest groups.

Public knowledge of data regulation will create trust and encourage participation in the digital economy. To this end, the government should be seen to be monitoring potential harmful dominance by key providers.

[5] https://www.scl.org/articles/10662-interoperability-a-solution-to-regulating-ai-and-social-media-platforms

[6] Cobbe, Jennifer and Veale, Michael and Singh, Jatinder, Understanding Accountability in Algorithmic Supply Chains (April 7, 2023). 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)

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?

    As the discussion paper notes, country responses have varied. We see that the initial US response involves minimal regulation so as not to constrain AI's potential for economic benefit and societal transformation, while the EU response favours a focus on human rights. We expect that the EU approach will come to dominate. Banning high-risk activities would bring us into line with the majority of the OECD.

RISK-BASED APPROACHES

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

    Yes. We support the approach of categorising AI types and requirements by risk levels and requirements (as extracted in Attachment B of the discussion paper). The EU 'AI Act' is by no means the only sensible approach, as we outline in our answer to Question 5, and note particularly the CoE Framework Convention on Artificial Intelligence,.

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

    Human rights based approaches are also an important consideration where fundamental rights of the citizen are at risk, as laid out in the draft CoE Framework Convention. We would be happy to provide more details on request.

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?

    Size could be a consideration to ensure regulatory compliance is not prohibitive to operations. The EU 'AI Act' makes a generic exception for non-public facing SMEs and this should be considered for Australia.

    Given the concerns around privacy, the ecosystem of LLMs is going to shift towards "not so large" language models (models in the order of 7-30B parameters compared with OpenAI 175+B models) that could be trained on private data and deployed on private low-cost infrastructures of these sectors.

    While larger companies like OpenAI are easier to be held accountable (due to their technical capability to conduct and improve safety aspects, potential brand damage risks, and the global legislative momentum), imposing the mandatory regulations on those privately built and deployed models is an area for consideration.

    Does the government intend to provide a framework with which these privately deployed systems should comply? For example, a financial institution which builds its own model might have satisfaction with certain aspects of the model performance that matters to them while missing several key safeguards. Meanwhile these companies do not have the expertise to evaluate all the various aspects of their models.

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

    We support the elements as presented and given the complexities, we recommend an expert working group be convened to research and establish the approach through extended research and more, targeted consultations.

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?

    There is already substantial policy and process development in this field. We refer the panel to an example by Dr Alessandro Mantelero (Council of Europe expert and leading academic Mantelero, A. (2022) Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI, Springer Press/T.M.C. Asser, at Beyond Data (oapen.org)

19. How might a risk-based approach apply to general purpose AI systems, such as large language models (LLMs) or multimodal foundation models (MFMs)?

    The EU 'AI Act' approach has been developed based on recent amendments to the draft legislation, notably at Article 28b. Monash researchers have intensively engaged with the European drafters of these amendments, in both Parliament and Commission, and can provide further details on request.

20. Should a risk-based approach for responsible AI be a voluntary or self-regulation tool or be mandated through regulation?

    It should be mandated through regulation. Self-regulation has been tried and failed numerous times over the past decade.

    In the United States and following consultation with other nations including Australia, the White House has obtained voluntary commitment from a group of seven large multinational AI leaders to standards relating to safety, security and trust[7]. While voluntary, it suggests a move to greater intervention to protect the public.

    And should it apply to:

    a) public or private organisations or both? Both

    b) developers or deployers or both? Both

[7] https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/