Submission in response to the Discussion Paper on *Safe* and *Responsible AI in Australia*

Lead author Dr Bernadette Hyland-Wood

Co-lead Responsible Data Science Program
Queensland University of Technology
Brisbane QLD Australia

E. b.hylandwood@qut.edu.au

Submission due date: 04 Aug 2023

Table of contents

About the QUT Centre for Data Science

The Queensland University of Technology (QUT) Centre for Data Science, based in Brisbane Australia, is a national and global leader in developing frontier methods for the use of data to benefit our world. We draw together capabilities in data science from across Australia to deliver world-class research, unique training opportunities, and ongoing engagement. QUT has made multimillion-dollar investments to fund core research on responsible data science, complex data analysis, human-centred AI, models and algorithms, and data for discovery. Through our work, the QUT Centre for Data Science bridges the gap between deep research and applications and is a trusted source for advice on data science to Queensland and the Commonwealth Governments. The QUT Centre for Data Science has a competitive advantage through its track record of cross-disciplinary projects, linkages with relevant Centres of Excellence and Cooperative Research Centres, and national and international networks.

Our interest in supporting safe & responsible AI

The QUT Centre for Data Science appreciates the opportunity to respond to the Department of Industry, Science and Resources' consultation for *Safe and Responsible AI in Australia.* We believe that artificial intelligence (AI) represents and has contributed to remarkable achievements in areas directly benefiting Australian society and our economy, and Australia's position within the international order. The transformative capabilities of data-driven systems integrating AI has revolutionised business operations and delivery systems through their ability to consume and process massive amounts of disparate structured and unstructured data. With the general availability of tools such as ChatGPT, among other competitive products, the capacity of AI to emulate human-like conversation, generate ideas, propose solutions and learn from patterns to generative novel outputs is rapidly making it an indispensable tool for knowledge production and informed decision-making in government, healthcare, life sciences research, financial services, mining, agriculture, and education.

The following recommendations are presented, and if considered and acted on, may allow civil society to leverage the vast potential of AI while mitigating potential risks. We offer guidance and provide actionable recommendations to increase multistakeholder engagement, nationally and internationally.

Executive Summary

Artificial intelligence (AI) also holds vast implications that are, as of yet, not fully understood by those who use it and are, directly and indirectly, impacted by AI. As multijurisdictional government organisations contemplate policies and legislation to guide and govern the availability and use of AI, there are numerous complex and nuanced issues related to the responsible use of AI within Australian industry, education, media and communication, among other sectors. Environmental, social and governance (ESG) considerations, stakeholder expectations, privacy issues, the future of Australia's workforce, security and defence, each require fit-for-purpose governance frameworks. This guidance, and focused regulation, will enable the public service and civil society to realise the opportunities and protect against unintended harms presented by AI.

It is uncontroversial that AI is one of the most critical technologies of our time. AI holds enormous promise while having far-reaching implications for our society, economy, security and defence. AI will continue to shape our day-to-day communication, commerce, government service delivery, foreign policy and security of the nation.

At the same time, current and emerging AI (super-intelligent AI) is an issue that is unprecedented due to the speed and scale of digital information sharing, centralisation of digital platforms, massive investments by the U.S. and China, and ultimately, the potential harms to society, economies and defence. No one understands how modern AI systems work. Period. They are massive inscrutable matrices of floating point numbers that software engineers 'nudge in the direction of better performance until they inexplicably start working' as intended, according to U.S. AI researcher and author Eliezer Yudkowsky.

Yet, until recently, Australia had not seen the speed, scale and ease of use of generative AI that is now widely available. **The ubiquity of these tools has clear and present dangers that must not be merely detailed in government reports, but urgently reviewed, considered and acted on within the next 12-24 months.** The practice of two-year consultations and reporting cycles will not be sufficient for the gravity of the issues facing Australia and our Pacific neighbours. Australia is highly reliant on digital platforms and cloud-based and dedicated data centres run by U.S.-based companies. The urgency with which the Australian Government, its allies and partners, must address safe and responsible AI is unprecedented.

Leading agencies within the Commonwealth Government, including the Department of Industry, Science and Resources, cannot afford to operate in business-as-usual (BAU) mode. Due to the unprecedented speed and ubiquitous integration of AI technologies into everyday information technologies used by government and civil society, urgent engagement between multijurisdictional government agencies, civil society and industry stakeholders is required.

Failure to engage substantively threatens the nation's economic growth, high-quality jobs, effective public health delivery, Australia's energy transition, and resilient defence and national security capabilities.

The policy areas that the Australian Government can materially influence the safe and responsible use of AI are:

- Multistakeholder governance at national and international levels
- AI ethics, governance and regulation of Australian entities
- Attracting and retaining data and AI literate public service and practitioners
- Policy settings for digital platforms used by the public service, industry and academia
- Workforce development capacity
- Compute power capacity in support of technology sovereignty

We offer recommendations that are intentionally not technical but rather focused on data and digital leadership *culture* that is required to address safe and responsible AI in Australia.

Recommendations

As governments around the world face accelerating and more severe climate events, pandemics, and the erosion of public trust fuelled by mis- and disinformation, public, incomplete or deficient national data strategies may leave public institutions unprepared for the strategic and tactical challenges they face. The countries that can most rapidly and effectively integrate AI into decision-making processes, both within civil society and defence settings, may be best positioned to navigate the challenges, embrace the opportunities and successfully strike the balance that maximises the opportunities posed by AI while ensuring it is used responsibly and avoids harms.

1. *Respond urgently* and refuse a '*business as usual*' culture. AI poses the potential for complex issues related to civil society's trust in institutions, privacy, and cybersecurity in ways not previously seen due to the scale and speed with which digital content travels.

2. Avoid the temptation to address the safe and responsible use of AI through central or large service delivery agencies. Leadership and funding should be directed to agencies with a proven track record with data standards, data governance, and methodology development and engagement with industry. These government organisations with relevant capacities are the scientific and statistical agencies who have evolved an understanding of data stewardship and data and digital culture. States such as New South Wales have also demonstrated exceptional innovation with *regulation as code.*

3. Engage and support *existing consortia and communities of practice* who possess expertise in data governance, data communities (i.e., surveillance & monitoring for public health), manufacturing and technology sovereignty.

4. Work with Australia's innovative technology and leading research communities (academic and public health institutes) on demonstrated approaches to improve data and digital literacy, data-driven systems and the societal impacts, data stewardship and cybersecurity.

The Geopolitical Landscape of Digital Platforms

Data-driven systems and the use of AI are not new. Australia has received immeasurable benefits for decades from vital AI-driven programs such as the U.S. Global Positioning System (GPS) and Himawari 9, the Japanese weather satellite system. Yet, evidence shows that data programs, and now the availability of AI faces immature and often contradictory policies, lack of technical proficiency and insufficient consideration of privacy and security considerations. This requires our immediate and renewed attention.

During the last decade, the trend has been for large technology companies to acquire or merge smaller companies to expand their capabilities, user base, and market reach. In the **U.S.A.** and **China**, there have been numerous notable instances of such consolidation in the technology sector, particularly with respect to data-driven systems. Digital platform companies in the U.S.A, including major technology firms including **Google**, **Microsoft**, **Facebook** (now Meta Platforms Inc), **Amazon**, and **Apple** have been acquiring small companies with expertise in data analytics, AI, machine learning (ML), and related forms of data-driven technologies for the last 15 years. These acquisitions are often aimed at enhancing a company's existing products or services, gaining a highly skilled workforce, or gaining a competitive market advantage. There are examples of predatory acquisitions aimed at protecting the market position of entrenched vendors, often in high-value sectors and markets such as government services. Some of these acquisitions have led to Australian-based highly skilled talent working for U.S.-based technology companies, particularly on distributed ledger technologies such as blockchain.

Australian governments and civil society depend entirely on foreign-owned and operated data and digital platforms. Due to effective localisation and pervasive marketing campaigns, Australians do not recognise foreign dependence on core digital technologies used within civil society and government (Australian Defence is the exception).

Data aggregation and analytics firms have emerged in the U.S. that specialise in aggregating and analysing vast amounts of data from various sources. Many of these firms have been venture-funded or received corporate investment, allowing them to grow significantly in size and influence, becoming targets for larger companies seeking to expand their data capabilities.

In China, the **Alibaba Group**, **Baidu**, **Huawei**, and **Tencent** have been prominent behemoths at the forefront of technology consolidation activities. These companies have extensive holdings in various sectors ranging from e-commerce, finance, entertainment, search, cloud infrastructure, and AI. **Alibaba Group, Baidu** and **Tencent** have focused on acquiring and investing in startups and established companies to broaden their respective portfolios of data-driven products and services. Huawei is popular throughout the Global South moving from a telecom supplier to a full-service digital transformation partner providing products and services. Chinese firm **ByteDance**, founded in 2012 by Zhang Yiming, owns **TikTok**, one of the world's fastest-growing social media platforms globally and is known for short-form videos. Central to

TikTok's success is uptake in Indonesia, Brazil, and West Asia. **ByteDance's** portfolio has expanded to include various content platforms and applications using AI. **Alibaba**, **Tencent**, **Baidu**, and **ByteDance**, along with some smaller companies have invested several billion dollars in tech startups in India, Singapore, Malaysia, and Indonesia.

The Chinese Government is investing 10x more than the U.S. (in computer vision AI). China is a fast follower in LLM R&D. The global reach of Chinese applications and the Chinese Internet imposes values, norms and legal frameworks on digital spaces that are used by Australian-based companies, governments, non-profits and research organisations. The full implications and impacts of this are emerging and not fully understood. The Chinese Government's policies and regulations have played a material role in shaping the consolidation of data-driven technology companies. Certain industries and sectors have experienced significant consolidation due to Chinese Government directives for strategic growth and ensuring data security and control. China has arguably amplified frictions by introducing a "Made-in-China" policy which is affecting how digital platforms are perceived and used outside of China. For example, India became the first country to ban and block WeChat and TikTok, along with many other Chinese apps, stating that they 'violated Indian sovereignty and security.'

The U.S.A. and China have both invested hundreds of billions over the last decade in data-driven systems. The Chinese Government has outpaced investment by any other country in data-driven defence systems. The consolidation of data-driven systems by two countries can and should raise concerns related to **competition, data privacy, governance, data sovereignty, and potential monopolistic practices**. The consolidation of digital technology platforms has not gone unnoticed by many governments who are scrutinising large companies' activities and are actively considering new regulations to address these issues.

It is worth remarking that due to strict Internet censorship laws and regulations, several countries, including China, Russia, North Korea, Iran, Cuba and Syria are *unable* to access ChatGPT and other U.S. based LLM technologies.

Australia's Position within the Global AI Landscape

From an international perspective, AI is shaping geopolitics, and through that, our society and economy.

> "Everything Australia wants to achieve as a country depends on its capacity to understand the world outside and to respond effectively to it."
> - Allan Gyngell, AO FAIIA

The Australian Government's Digital Economy Strategy has published an ambitious vision for Australia to have a robust digital economy and society by 2030. Artificial intelligence-powered digital platforms and automated decision-making (ADM) present significant opportunities and

threats in advance of 2030. The issues papers *Positioning Australia as a Leader in digital economy regulation - Automated decision making and AI Regulation* (2022) and *Safe and Responsible AI in Australia* (2023) provide useful definitions and offer in-depth discussion on large language models (LLMs) and multimodal foundation models (MFMs). However, the aforementioned reports and initiatives occurred and were published prior to the November 2022 release of OpenAI's ChatGPT.

Until recently, Australia had not seen the speed, scale and ease of use of generative AI that is, until November 2022. OpenAI's release of ChatGPT 3 disrupted the landscape. The ChatGPT Web service gained 1M users in the first week of its release. As of February 2023, the uptake of the LLM-based conversational application increased exponentially to 100M users. The U.S.-funded technology firm **OpenAI** is set to exceed US$200M in 2023 revenue. Many fast followers and tens of billions of US dollars are being invested in LLMs and multi-modal foundation model (MFM) technologies.

Therefore, we must be realistic. Australia can only dream of the the primacy that the U.S. and China have and will continue to have in AI research and development. And while Australia is home to advanced research institutions and leading researchers, many of our post-graduate students leave and work for foreign companies. The phenomenon of highly-skilled data scientists and engineers moving overseas or remaining in situ, often on lifestyle regions such as the Sunshine Coast and Gold Coast (QLD) while employed by overseas companies is a well-known phenomenon within the tech community. Australia relinquished its reputation as an innovator and fast adopter of digital technologies for various reasons that are well-documented in many industry reports and white papers.

While the Safe and Responsible AI in Australia Discussion Paper aims to cast Australia with a positive glow, the reality for experienced international tech entrepreneurs (including the author) is that Australia's graduate and tech innovators are warmly embraced and well-remunerated by primarily two AI leaders, the U.S. and China.

Without diminishing the Australian Government's progress through the eSafety Commission's progress on safeguarding Australian citizens online and proposal for prescriptive AI Ethics Principles, these guidelines are helpful, but not widely socialised. Pragmatic, culturally instilled and continuously reinforced commitment to data literacy, cybersecurity practice and data stewardship are required.

The practice of two-year consultations and reporting cycles will not be sufficient for the gravity of the issues facing Australia and our Pacific neighbours. The Australian Government is advised not to be content that we have world-leading research AI capacities yet address it with the same risk-averse approach that has been the anemic national data strategy.

The pervasiveness and ease of use of current LLM tools presents a clear and present danger that glossy reports will not address. The safe and responsible use of AI is a

strategic and tactical priority that must be funded and acted on within the next 12-24 months.

Consultation questions and responses

Definitions

Question #1

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

Yes, the key definitions and sources cited in Figure 1 on Technologies and Applications are accurate and sourced from reputable, well-regarded sources.

Potential gaps in approaches

Question #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?

Australia's data market has been essentially unregulated for more than 40 years. The data is on citizens and often the government collected much of it. Personal data gleaned from our customer loyalty programs, air travel, and social media behaviour, is all available to 3rd parties. Digital platform vendors including Facebook and Google are large customers of the same data as the AdTech ecosystem. These data are likely to have been used to train the Foundation Models and LLMs powering recent digital platform services. Regulation is required to reduce the unintended harms of these data collection and derivative products and services.

Potential harms from unregulated AI are far-reaching. Harms and unintended consequences include but are not limited to:

> ***Misinformation and deep fakes:*** Generative AI can be used to create highly realistic fake content, including deep fakes videos and articles that can spread mis- and disinformation at scale, at the speed of the Internet. Considerable scholarship has been dedicated to the significant challenges to trust in media, public discourse, and political stability (Bruns 2023; Ha, Graham & Grey 2022; Bruns, Schumacher, et al. 2022; Angus, Bruns, Hurcombe, et al 2021)

Privacy and Data Protection: Generative AI models are often trained on large datasets that may contain personally identifiable information (PII), private or sensitive information, derivative works, protected or sacred data (i.e., Indigenous data not authorised by the data steward).

***Bias and discrimination*:** When Large Language Models (LLMs) are trained on datasets, there is inevitable bias in the data. : Biased data may perpetuate and amplify social biases, and deficit narratives, rapidly resulting in discriminatory or unfair recommendations or outputs. Governments as a major consumer of data-driven platforms need to proactively address biases and unintended consequences in order to treat all citizens in a just and equitable manner.

Other harms and unintended consequences addressed in numerous academic and industry collaborations include **intellectual property (IP) infringement, job displacement** and **manipulation and persuasion at scale.** The latter is addressed in-depth by the submission from the **The ARC Centre of Excellence for Automated Decision-Making and Society (ADM+S).**

Question #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.

Given the technology sector's consolidation and control by two countries (the U.S.A. and China), the **practical approach for the Australian Government** is to actively **support and engage in multistakeholder governance**, in collaboration with industry and academics, through international research consortia, communities of practice and international organisations including the **World Wide Web Consortium, Transparency International, United Nations Open SDG Data Hub, Research Data Alliance, AsiaPac Regional Internet Governance Forum**). In Australia, there are communities of practice including the **Australian Government Linked Data Working Group** and **Indigenous Data Governance** Initiative (based out of the University of Melbourne). There are several established organisations that operate on significant goodwill and often modest funding. They have supported world-leading research, standards and best practices for data-driven systems, transparency, privacy, and Indigenous governance.

Question #4

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

See response to Question #3.

Responses suitable for Australia

Question #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?

No liberal democratic government, including the United States and European Union, yet has a **comprehensive regulatory framework to manage generative AI.** The U.S. Government has stepped up engagement with industry through the White House Office of Science and Technology Policy (OSTP) with the ***Blueprint for an AI Bill of Rights*** under the current administration. On the momentum of the internationally respected GDPR legislation, the European Union has continued to work diligently and collaboratively on trustworthy AI. The EU published its proposed ***Artificial Intelligence Act (2023)***. The U.S. and EU have vital executive support to define frameworks for the human-centric and ethical development of AI, and transparency and risk-management rules for AI deployed in the U.S. and Europe, respectively.[1] The de facto standards and legislation are likely to be widely adopted by other nations. However, they will remain high-level and prescriptive and a minimum of five or more years to 'socialise' within industry, government and academia based on prior disruptive digital technologies.

In the European Union, under the EU AI Act, governance systems based on risk management are required to meet certain standards set forth relating to:

- Data quality and relevance (including examining data for biases)
- Technical documentation
- Record keeping (logging etc)
- Transparency and provision of information to users
- Human oversight
- Accuracy, robustness and cybersecurity
- Specific additional standards/requirements have been proposed for foundation models

Unlike product safety standards, 'fitness for purpose' has long been recognised and remains unresolved when applied to data and software. Standards bodies are numerous and often domain-specific, resulting in no consensus on data exchange standards, best practices, privacy, and safety considerations, outside of specific industries (e.g., finance) and are typically for specific purposes, such as recordkeeping and regulatory reporting.

---

[1] U.S. *Blueprint for an AI Bill of Rights, see* https://www.whitehouse.gov/ostp/ai-bill-of-rights/ The European Union recently published its proposed *Artificial Intelligence Act (2023)*, https://artificialintelligenceact.eu/

Target areas

Question #6

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

Responses within the public and private sectors by necessity will be very different. While the industry awaits policies and regulations, they will respond to pressing international trade and operations requirements through adoption of practices including general liability principles that are sector and industry-specific and aim to avoid harms (i.e., personal safely, financial, legal).

Question #7

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

The Australian Government is encouraged to engage with research experts, industry stakeholders, and Indigenous representatives to take on board recommendations and public sentiment on the use of AI.

Additionally, master classes for SES and EL-level staff on data governance, generative AI and key principles is required. Data fluency and a mature understanding of data stewardship in 2023 and beyond must be a mandatory skill set for government leadership. Data and digital cannot be outsourced to consultants.

A relevant source that details the deficiencies that permeate the APS is the Productivity Commission's 2017 report on Data Availability and Use. Substantial research points to the necessary practice and cultural changes that are required across government agencies.

Question #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.

Avoid technology-specific regulation. Instead, focus on activities, behaviours and harms that technologies pose.

Question #9

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

> *a) Where and when will transparency 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.

No response.

Question #10

Do you have suggestions for:

a) Whether any high-risk AI applications or technologies should be banned completely?

Yes, in keeping with international consensus within liberal democracies, lethal drones with facial recognition, armed robots, and autonomous fighter jets should be banned completely.

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

No response. The author does not have expertise in this area.

Question #11

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

Implications and infrastructure

Question #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?

No response.

Question #13

What changes (if any) to Australian conformity infrastructure might be required to support assurance processes to mitigate potential AI risks?
No response.

Risk-based approaches

Question #14

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

No response.

Question #15

What are the main benefits or limitations of a risk-based approach? How can any limitations be overcome?

No response.

Question #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?

No response.

Question #17

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

No response.

Question #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?

No response.

Question #19

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

No response.

Question #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:

- public or private organisations or both?
- developers or deployers or both?

No response.

References

Angus, D., Bruns, A., Hurcombe, E., Harrington, S., Glazunova, S., Montaña-Niño, S. X., … Dehghan, E. (2021). 'FAKE NEWS' AND OTHER PROBLEMATIC INFORMATION: STUDYING DISSEMINATION AND DISCOURSE PATTERNS. AoIR Selected Papers of Internet Research, 2021. https://doi.org/10.5210/spir.v2021i0.12089

Bell, G., Burgess, J., Thomas, J., and Sadiq, S. (2023, March 24). Rapid Response Information Report: Generative AI - language models (LLMs) and multimodal foundation models (MFMs). Australian Council of Learned Academies

Bernshteyn, R., (2020). Smarter Together. How Communities Are Shaping the Next Revolution in Business, Greenleaf Publishing

Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Liang, P. (2021). On the opportunities and risks of foundation models. arXiv prepint arXis:2108.07258.
Srivastava, N. 2023 Friend or foe? Psychbots and the law, LSJ online, https://lsj.com.au/articles/friend-or-foe-psychbots-and-the-law/

Helberg, J., (2021). The Wires of War: Technology and the Global Struggle for Power. Simon and Shuster.

Hyland-Wood, B. (2021). Bridging the Open Data and Public Policy Gap: Barriers and facilitators for effective government data sharing

Yudkowsky, E. (2008). Artificial intelligence as a positive and negative factor in global risk. Global catastrophic risks, 1(303), 184.

Zuiderwijk, A., Chen, Y.C., Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and research agenda. Government Information Quarterly, 38(3) 101577.