Kingston AI Group
Australian Professors of Artificial Intelligence

kingstonAIgroup.org.au

Safe and responsible AI in Australia

Submission by the **Kingston AI Group:** *Australian Professors of Artificial Intelligence*

26 July 2023

The Kingston AI Group comprises 13 leading professors from eight universities and the Chair of Robotics Australia Group. Eight are members of the learned academies and five are existing or previous laureate fellows. As well as driving Australia's artificial intelligence (AI) research, the group's members are working with companies to develop commercial AI solutions.

AI is a transformational technology that is impacting every sector of industry and society, solving previously intractable problems and delivering productivity gains across the economy. In fact, there is no other technology with the potential to improve Australia's productivity as broadly as AI.

Quite rightly, the government is considering how the technology should be regulated in Australia. How Australia approaches the regulation of AI is an important discussion on how best to protect the interests of Australia and its citizens. Moreover, a more certain regulatory environment supports public assurance in the adoption of the technology and provides a greater degree of confidence for businesses to invest, innovate and compete in international markets.

Safe, ethical and responsible AI in Australia requires investment in research and development

To ensure the safe, ethical and responsible deployment of AI, it is critical that Australia lifts its investment in its capacity to create, understand and control the technology—this cannot be achieved by regulation alone. In fact, the OECD's five primary policy principles on which regulation and development of AI technologies should be based (and to which Australia has committed), specifically include investment in AI research and development[1]. The OECD recognises that R&D represents a key mechanism for the creation of more capable and trustworthy systems.

Many comparable nations are investing heavily in their AI ecosystems, including R&D, as they recognise the advantages to their economies and societies.

For example, the United Kingdom invested an initial £1 billion (≈A$1.87 billion) from 2018, in a number of initiatives including a significant contribution to R&D and measures to support the growth of the broader AI ecosystem[2]. Earlier in 2023 a further £250 million (≈A$445 million) investment was made for technology missions in AI, quantum technologies and engineering biology as part of a larger boost to innovation with the objective of setting the country up as a technology superpower by 2030[3]. In June 2023, the UK Prime Minister announced a £100 million (≈A$190 million) investment in an AI safety

[1] *OECD AI Principles*. Adopted May 2019
[2] (2022) *National AI Strategy*. Department for Science, Innovation and technology. Government of the United Kingdom
[3] *Plan to forge a better Britain through science and technology unveiled*. Press release 6 March 2023

taskforce, which includes an allocation for research[4]. The UK recognises that the safe and responsible deployment of AI is underpinned by a solid foundation of R&D.

The Singapore government is also investing heavily in AI R&D (≈A$740 million announced in 2019) with an emphasis on not simply adopting new technology, but rethinking new business models, impactful productivity gains and new growth areas. Moreover, their national AI strategy addresses areas where attention is needed to manage change and/or manage new forms of risks that emerge when AI becomes more pervasive[5]. This is also the experience in France, Canada and Germany—all are making large investments in technology creation through R&D, and all see these investments as central to safe, ethical and responsible AI.

However, to date, Australia has only offered piecemeal, modest measures to support AI, while other nations accelerate ahead. This represents a sovereign risk for Australia as we become increasingly dependent on technology created elsewhere and subject to regulatory regimes over which Australia has little to no influence. Australia must lift its investment across the breadth of the AI innovation value chain on a solid foundation of R&D.

To put it simply, a strong domestic AI research and development capability is crucial to achieving responsible AI. If Australia's most valuable industries, natural resources, and national datasets, are instead subject solely to the diffusion of AI from international providers—where any benefits are parasitised by international shareholders—that would deliver an alarming scenario of *irresponsible* AI.

Australia should adopt a balanced risk-based approach to regulation

Australia should adopt a balanced approach to AI regulation that protects the interests of the community, but does not stifle innovation in a critical field with the potential to deliver broad benefits across the economy and society.

While a balanced approach to regulating certain applications of AI is warranted, we don't support regulating AI research in addition to the usual expectations of academic research generally, such as the Australian code for the responsible conduct of research or research domain-specific requirements, such as in medical research. The potential implications of AI research are so broad-based that regulating AI research specifically may simply serve to stall innovation without addressing legitimate issues of concern. Moreover, we risk limiting Australia's ability to develop a sovereign capability in technology that will nonetheless impact society.

AI on its own is very rarely the problem, rather it's the inappropriate *use* of the technology that can be problematic. Therefore, it's the uses of AI that should be the target of regulation. As described in a recent report commissioned by Australia's National Science and Technology Council, four of the six major steps in AI product development (as distinct from research) occur before the public release of a product[6], so we recommend that regulation which is proportionate to risk should apply to the product design and development process and not just once a product is available to consumers; particularly in relation to data capture, storage and use.

We support a flexible risk-based approach that is proportionate to the risk of failure or misuse of technology in certain contexts. In particular, any mandatory regulation should be reserved for the highest risk applications and be focused on identifying where the responsibility for failure and/or misuse lies,

[4] Morrison, R. (2023) *Rishi Sunak touts UK AI safety research at London Tech Week*. TechMonitor
[5] (2019) *National Artificial Intelligence Strategy: Advancing our Smart Nation*. Smart Nation Digital Government Office.
[6] Bell, G., Burgess, J., Thomas, Thomas, J. and Sadiq, S. (2023) *Rapid Response Information Report: Generative AI-language models (LLMs) and multimodal foundation models (MFMs)*.

including protection of the data used to train the systems. In these cases, it should be recognised that many AI technologies have a diversity of applications, and as such, regulation should avoid negative unintended consequences. For example, large-scale facial recognition for surveillance is highly problematic if used in an unregulated way, but facial recognition is also commonly used to organise family photos on smartphones. Moreover, it is important to acknowledge that the extent to which a specific application of AI may warrant regulation may not be immediately obvious at release with quite lagged negative impacts.

As a relatively small, open market economy, Australia must harmonise its regulatory environment as best as possible with those of key international partners and be consistent with the OECD policy principles. Australia's regulatory framework should be sufficiently flexible to adapt to advancements in the technology and changing international circumstances, particularly given the dynamic nature of the international regulatory landscape. It's important Australia doesn't simply adopt the most restrictive position as this may unnecessarily stifle the sort of innovation that may serve to make AI safer and more trustworthy, and limit Australia's capacity to harness the technology to the nation's economic and social benefit.

We thank the government for the opportunity to provide a submission to the *Safe and Responsible AI in Australia* discussion paper. We have provided responses to a number of specific questions in the paper where we believe we are well placed to make an informed contribution.

In summary, the Kingston AI Group makes the following recommendations.

Recommendation 1: Substantially and urgently increase investment in Australia's AI research and development to build a sovereign capability that is essential to ensure safe and responsible AI in Australia.

Recommendation 2: Ensure a flexible and risk-based approach to AI regulation that is consistent with the OECD policy principles and adaptable to future developments.

Definitions

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

It is unnecessary to include the term “predictive” in the definition of AI (Figure 1). Generative methods are not necessarily predicting anything. The definition also does not appear to capture emergent behaviour that is not one of the given “human-defined objectives”, yet these emergent properties seem to be what applications are built on and rely upon, and thus arguably the most important considerations when it comes to safe and responsible AI.

The definition of machine learning should include the ability of the machine / agent to improve with experience, which may or may not be captured by the term “training data”.

We would avoid anthropomorphising characteristics of AI. For example, we would recommend avoiding the term “hallucination” (Section 2.2). In some sense all generative models are hallucinating and the only models that don’t are retrieval models. We should instead stick to well defined terms such as “prediction” or “estimation.”

Box 2 talks about “being safe” as something guaranteed by Australian consumer law. But it’s unclear what this means in the context of AI. Something that is safe for one user may not be safe for others. For example, motor vehicles are usually safe when operated by trained drivers, but dangerous when operated irresponsibly.

Box 3 should apply to all technology, AI being a small subset.

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?

Section 2.2 raises algorithmic bias as being one of the biggest risks of AI. We would note that algorithmic bias is not unique to AI, as it’s a property of any decision-making process. Perhaps a more significant risk posed by AI is the intentional misuse and concentration of control of a powerful, unregulated technology in the hands of foreign corporations and outside Australian jurisdiction.

Regulation needs to include public disclosure of what data algorithms are trained on, and assurances of data sovereignty and privacy. Moreover, an important regulatory tool would be to ensure audit capability. For example, any company deploying AI technologies in Australia might be required to allow audits of its products, training algorithms and datasets by an independent expert Australian authority, which would include academic representation.

There is also a risk of a significant technical debt for products that build on foundational AI infrastructure such as large language models. GPT3.5 and GPT4 have qualitatively different behaviour. A product or service that is built on GPT3.5 will not work the same when switched to GPT4. One potential area for regulation is therefore to require basic AI infrastructure providers to guarantee backwards compatibility and announce any deprecation of features / versions with sufficient time for downstream users to respond.

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.

For Australia to have a meaningful impact on the responsible use and development of AI, both domestically and abroad, we must invest in our own AI capability, including R&D, education and training.

A real risk for Australia is that by failing to keep pace with comparable nations, we'll instead be largely subject to the regulatory arrangements of the jurisdictions where the technology is created.

The Australian Government's investment in the National AI Centre's Responsible AI Network is a positive initiative to support industry in Australia to adopt responsible AI practices in consultation with experts / industry leaders from around the country.

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.

Any regulatory body or organisation / initiative aimed at enhancing, coordinating or controlling AI deployment and development should be done by an independent statutorily established entity and ideally not through an existing organisation such as the CSIRO.

There is much to learn from the UK example, where there is an AI Council, Foundation Model Taskforce, and Office for Artificial Intelligence. Each plays a different role.

Australia could have an expanded AI council that advises government on how to ensure the country gains the most value from the AI revolution. It would comprise experts in AI technology from academia and industry, business leaders, defence leaders and successful AI entrepreneurs. This would also be consistent with a recommendation of the Australian Human Rights Commission[7].

We propose an Office for Artificial Intelligence that reports to the Department of the Prime Minister and Cabinet on the uptake of AI across the Australian Government and the economy. The office should collect data and undertake cross-departmental planning, and be responsible for monitoring any AI-specific regulations, and general regulations that pertain to AI.

We propose a national, multi-agency research institute that focuses on researching, developing and testing responsible AI, with the aim of becoming the *'brain'* for the Office for Artificial Intelligence, developing AI tools for the national benefit, monitoring and auditing systems for government agencies. Such an institute could help address the skills gap in AI in Australia by funding PhD, master degree and short course training scholarships in safe and responsible AI.

We back the National AI Centre playing its role in supporting existing businesses to adopt AI and building understanding and awareness of AI across the country.

[7] Farthing, S., Howell, J., Lecchi, K., Paleologos, Z., Saintilan, P. and Santow, E. (2021) *Human Rights and Technology: Final Report*. Australian Human Rights Commission.

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?

The Global Partnerships in AI (GPAI), of which Australia is a strong supporter and contributor, is developing a range of practical tools for AI and data governance. Australia should take advantage of the work that is being done by GPAI and seek out tools that match the requirements of its final policy objectives and systems. France and Israel have recently released major strategies to ensure that they have access to the in-country AI talent they will need for the future[8,9].

The United States is investing US$200 billion into research into AI, robotics, quantum computers and related technologies[10]. This would be approximately US$13.3 billion (A$19.7 billion) if adjusted for GDP and applied to an economy the size of Australia's.

There are dozens of similar national strategies. Without the talent, AI regulation and policy will be much less effective. Australia should invest substantially to increase the capabilities of its workforce, at all levels of training.

Target areas

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

The public sector should model best practice in how it applies AI in Australia's interests.

There are particular national datasets that are held and managed by the public sector that belong to the Australian public, and are repositories of great commercial and social value. Because of their value, multinational companies will lobby for governments to make these datasets available to enable the creation of supposedly 'public good' AI tools. However, this will create the potential for multinational companies knowing more about Australia and Australians than the Australian public.

In areas such as public health, national security, taxation, and valuable sovereign assets (minerals, fisheries etc.), governments should exercise a high degree of caution in making datasets available to the private sector. There is a strong argument that it's in the national interest for governments to work with the university sector, CSIRO, and other public institutions, to translate these datasets into domestic AI assets that can benefit all Australians.

There is an opportunity to establish a national capability to develop new AI tools that can transform the productivity of the public sector. There seem to be few other options other than productivity growth for government to be able to maintain high quality services without an ever-higher tax burden on Australians. There are major opportunities that could be seized right now in public health, the NDIS, social security, defence, national security, and energy transition, that could achieve major improvements in efficiency with the development and adoption of high-quality AI tools by trusted agents.

Science itself is being transformed by AI. Improvements in efficiency of 10X, and even 1,000X are being achieved. For example, Google's AlphaFold has transformed the ability for scientists to discover new

[8] *Choose France – CNRS AI Rising Talents*. Press release 19 June 2023
[9] Kogosowski M. (2022) *Israel approves national plan for increasing, developing human capital in high-tech*. Israel Defense.
[10] (2022) *What the CHIPS and Science Act means for artificial intelligence*. Stanford University: Human-Centred Artificial Intelligence Explainer

drugs and predict the behaviour of protein structures. The returns to the taxpayer of publicly funded research to develop the next generation of AI tools to transform science would be immense.

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

There are first some basic mechanical challenges to overcome if government agencies are going to be able to develop their own AI. This is the surest path to safe and responsible AI because the agencies establish the standards and requirements and not instead try and retrofit someone else's design principles into their own.

In the current paradigm of 'big data AI', departments may be bound to the terms of the IT supplier agreement to only use particular servers and data transfer protocols that may preclude experimentation on government data to test new, potentially transformational tools. In addition, many agency data management and governance systems are not fit for purpose in an AI world, and it can take many years to negotiate access to data from different government units to create a dataset that can be used to train powerful AI systems that could transform public sector efficiency. Moreover, data and training paradigms may change radically in future, necessitating further substantive revisions to these processes.

Government procurement rules limit the ability of government agencies to test and 'play' with research and innovative companies to develop and pilot AI systems that could then be adopted. The result is that governments will only purchase pre-developed systems that are rarely fit for purpose and may not meet government requirements for safety, privacy and responsibility. The power of government procurement in supporting the translation of university research was recognised in the Australian Government's *Australian Universities Accord – Interim Report* as an area for further consideration.

Leaving aside these mechanical issues, governments have already set standards for safe and responsible AI, but government agencies will find it difficult to find AI experts to enable the translation of those principles into systems and policies. There is major deficit in AI expertise that needs to be addressed as a matter of urgency. In particular, Australia produces insufficient numbers of PhD graduates in AI for the public service, let along the rest of the economy. The is a strong case to be made for government agencies to team up with universities to directly fund internships, master degree scholarships and PhD scholarships to help overcome the deficit in skills and test and experiment to build organisational capacity to fully engage in this potentially transformational technology.

There are also cultural challenges to integrating AI into the public service. Recent examples have shown that algorithmic processes, no matter how simple, can be badly misused if the culture of the public service is not focused on delivering lawful value to the public[11]. AI is a powerful tool, and like any powerful tool, can cause damage if used to exacerbate suffering and inequality.

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.

The risks associated with AI relate predominantly to how the technology is used in specific contexts. Therefore, a technology-specific approach is favoured in the vast majority of circumstances. However, there may be a few limited examples, such as the application of generative AI models, where more generic requirements concerning transparency about when such models should be used and the source of the data upon which they've been trained, may be meritorious.

[11] (2023) *Royal Commission into the Robodebt Scheme*

Governance is the key solution to addressing any risks from AI. Our experience is that AI developed and delivered in good faith usually works as intended, delivers great value and is quickly corrected where unforeseen issues arise. AI that is deliberately designed to deliver profit at the expense of people is difficult, time-consuming and costly to unpick. Australia has consumer protection laws but these will only be powerful if cases are monitored and resources are available to stop malevolent uses of AI.

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?

Watermarks are likely to play a large part in identifying content created using generative AI tools, as highlighted by the recent agreement reached by large tech companies including Amazon, Google, Meta, Microsoft and OpenAI, and the Biden administration[12]. We support such initiatives.

Public trust and confidence would also be supported by addressing auditability and verifiability of AI tools and products.

10) Do you have suggestions for:

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

We would be reluctant to suggest an outright ban on any AI technology, or any application, immediately. But we do advocate, here and elsewhere in this submission, for an application-based and risk-based regulatory environment which may, after careful and informed consideration, go so far as to impose a ban.

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

The government has a role to play in improving the broader public's AI literacy. Most people already use some form of AI every day but are not aware of doing so. The government has a role to play in educating the public about the benefits of AI, and resourcing education, training and innovation initiatives.

[12] Paul, K., Bhuiyan, J. and Rushe, D. (2023) Top tech firms commit to AI safeguards amid fears over pace of change. *The Guardian*.

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?

Outright bans on dual-use technology risk undermining investment in development of tools with broad application and benefit. However, in the case that outright bans are implemented, it should be acknowledged that Australians would nonetheless encounter technology created overseas and that Australian R&D related to these tools is essential to maintain a degree of understanding and control of the technology, as well as a sovereign capability that ensures we are not entirely reliant on other nations.

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

It is very unlikely that all risks can be characterised and mitigated ahead of time. Regulatory frameworks will need to be revisited over time to account for new shifts in technological capability that bring with them unexpected implications both positive and potentially negative. Moreover, some unforeseen negative implications may lag the initial deployment by considerable periods. Therefore, while a risk-based approach is warranted, it must be acknowledged that there will very likely be impacts that will need to be addressed after the fact. One such example could include less apparent risks (such as those social media have had on the welfare of young people). They're likely too low risk to be captured by a risk-based approach. But over the longer term, and when applied to large populations, the impacts can nevertheless be significant.

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:

Self-regulation may be appropriate on a case-by-case basis for lower risk activities, while higher-risk activities warrant mandated regulation.

Authors

Professor Joanna Batstone
Director of the Monash Data Futures Institute, Monash University

Professor Peter Corke FAA FTSE FIEEE
Joint Director of the QUT Centre for Robotics, Queensland University of Technology

Professor Stephen Gould
Australian National University

Professor Anton van den Hengel FTSE
Director of the Centre for Augmented Reasoning, The University of Adelaide

Adjunct Professor Sue Keay FTSE
Robotics Technology Lead, OZ Minerals; and Chair of the Robotics Group Australia

Professor Jie Lu AO FIEEE
Director of the Australian Artificial Intelligence Institute (AAII), University of Technology Sydney (UTS)

Professor Simon Lucey
Director of the Australian Institute for Machine Learning (AIML), The University of Adelaide

Professor Michael Milford FTSE
Joint Director of the QUT Centre for Robotics, Queensland University of Technology

Professor Ian Reid FAA FTSE
Australian Institute for Machine Learning (AIML), The University of Adelaide

Professor Ben Rubinstein
Associate Dean (Research) Faculty of Engineering and Information Technology, The University of Melbourne

Professor Svetha Venkatesh FAA FTSE
Co-Director of the Applied Artificial Intelligence Institute (A2I2), Deakin University

Professor Toby Walsh FAA
Chief Scientist of the UNSW Artificial Intelligence Institute, The University of New South Wales