Committee Secretary

Joint Committee of Public Accounts and Audit

PO Box 6021

Parliament House

Canberra ACT 2600

28th October 2024

Dear Officer,

RE: Inquiry into the use and governance of artificial intelligence systems by public sector entities

The Australian National University Law Reform and Social Justice Research Hub ('ANU LRSJ Research Hub') welcomes the opportunity to provide this submission to the Legal and Constitutional Affairs Reference Committee, responding to the terms of reference.

The ANU LRSJ Research Hub falls within the ANU College of Law's Law Reform and Social Justice program, which supports the integration of law reform and principles of social justice into teaching, research and study across the College. Members of the group are students of the ANU College of Law, who are engaged with a range of projects with the aim of exploring the law's complex role in society, and the part that lawyers play in using and improving law to promote both social justice and social stability.

Summary of Recommendations:

- 1. Transparency statements include the requirement to disclose the kinds of decisions made using AI;
- 2. That potential future uses of AI by public entities are investigated and any that pose an unacceptable risk are prohibited;
- 3. Transparency in AI systems, including a 'right to explanation' for AI based decisions;
- 4. That legislative protections are implemented for individuals to contest or seek human review of decisions made by AI;
- 5. That the use of 'black box' AI in the public sector is regulated such that it cannot be used without significant human oversight for high risk applications;

Inquiry into the use and governance of artificial intelligence systems by public sector entities Submission 31

- 6. That an independent oversight entity is created to review automated decision making in the public sector;
- 7. That the Committee recommend the Parliament create a regulatory body to monitor AI development and use by government agencies, develop AI regulation laws, and investigate unethical practice and tools.

If further information is required, please contact us at

On behalf of the ANU LRSJ Research Hub, Authors: Jamie Cheeseman, Daniel Marns, Mekala Navaratne, Ben Archibald Editors: Jae Brieffies

The purposes for which AI is currently being used by the public sector entity and whether there are planned or likely future uses

Artificial intelligence ('*AI*') can cover a range of technologies but broadly may be defined as software with the capacity to carry out tasks that usually require human intelligence.¹ The Organisation for Economic for Cooperation and Development ('*OECD*') defines AI as 'a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions content, recommendations, or decisions that can influence physical or virtual environments'.² It may include technologies including machine learning, neural networks, natural language processing, robotic process automation, computer vision and robotics.³ While AI may increase efficiency, thereby increasing capacity, it also comes with significant risks and ethical challenges.

Use of AI in the public sector has been increasing and is often justified by the goal of making the delivery of services by public sector entities more efficient and reliable.⁴ In the public sector, AI is currently being used in the form of chatbots, document and image processing, data mapping, text recognition and translation and natural language processing.⁵

There is specific legislation empowering government agencies to make decisions using computers in areas such as migration, social security, superannuation and health.⁶ For example, s 495A of the *Migration Act 1958* (Cth) empowers the Minister to arrange for the use of computer programs to make decisions, exercise power or comply with an obligation. Similarly, under s 6A of the *Social Security (Administration) Act* 1999 (Cth) the Secretary may arrange for computer programs to make decisions under social security law. The Minister or the Secretary is made responsible for the decision of the computer program, as if they had made the decision themselves.⁷ Decisions in the areas of migration and social security for example, have the potential to be hugely consequential in people's lives. It is therefore essential to have appropriate regulation and oversight where computer programs are involved in decision making in the public sector.

¹ Artificial Intelligence in the Public Sector (World Bank, 2020).

² OECD, *Explanatory Memorandum on the Updated OECD Definition of an AI System* (OECD, 5 March 2024) ³ 'Adoption of Artificial Intelligence in the Public Sector', *Australian Government Architecture* (Web Page) <<u>https://architecture.digital.gov.au/adoption-artificial-intelligence-public-sector-0</u>> ('*Adoption of Artificial Intelligence in the Public Sector*')

⁴ Ines Mergel et al, 'Implementing AI in the Public Sector' [2023] *Public Management Review* 1, 2.

⁵ 'Adoption of Artificial Intelligence in the Public Sector' (n 3).

⁶ See, eg, Migration Act 1958 (Cth) s 495A; Social Security (Administration) Act 1999 (Cth) s 6A; Superannuation (Government Co-contribution for Low Income Earners) Act 2003 (Cth) s 48; My Health Records Act 2012 (Cth) s 13A.

⁷ Migration Act 1958 (Cth) s 495A(2); Social Security (Administration) Act 1999 (Cth) s 6A(2).

Inquiry into the use and governance of artificial intelligence systems by public sector entities Submission 31

In March this year, a 6 month trial of Microsoft's Copilot commenced in over 50 public service agencies.⁸ The trial intends to explore how generative AI may assist with certain tasks such as meetings, correspondence and presentations, so that people can focus on more complex tasks.⁹ Five principles to guide the use of Microsoft Copilot are based on Australia's 8 AI Ethics Principles and participating staff are required to complete learning modules and knowledge assessments.¹⁰

The need for caution in utilising generative AI such as Microsoft Copilot is demonstrated by a case where a psychosocial safety training module generated using Microsoft Copilot used the name of a real sexual harassment complainant.¹¹ The trainer who had used the chatbot to generate the scenario did not know that the scenario was not fictional.¹² This example demonstrates the need for caution in utilising generative AI such as Microsoft Copilot and the need for comprehensive training for individuals even where the tasks given to the chatbot seem straightforward.

While generalised information about how AI may be used by public sector entities is available, clear and detailed disclosure of how AI is being used in the public sector could not be found. Under the new *Policy for Responsible Use of AI in Government*, departments and agencies must public transparency statements by 29 February 2025.¹³ These statements must include information about why the agency is using AI and classification of the AI used where the public may interact with or be significantly impacted by it.¹⁴ They must be written in clear plain language and provide a public contact email for more information.¹⁵ They also must be updated at least once a year or whenever a significant change is made.¹⁶ This is a good start to providing more transparency around the purposes for which AI is used by public sector entities. However it does not require departments and agencies to disclose specifically which decisions are made

⁸ 'APS Trials Generative AI to Explore Safe and Responsible Use Cases for Government', *Australian Government Digital Transformation Agency* (Web Page, 9 March 2024)

<https://www.dta.gov.au/blogs/aps-trials-generative-ai-explore-safe-and-responsible-use-cases-government> 9 Ibid.

¹⁰ Ibid.

¹¹ "Warning AI "hallucinates" and "Can't Be Relied on" after Chatbot Uses Real Information for "Fictional Scenario", *ABC News* (online, 20 August 2024)

https://www.abc.net.au/news/2024-08-21/ai-chatbot-psychosocial-training-bunbury-regional-prison/104230980.

¹² Ibid.

¹³Transparency Statements', *digital.gov.au* (Web Page)

<https://www.digital.gov.au/policy/ai/transparency-statements>.

¹⁴ Ibid.

¹⁵ Ibid.

¹⁶ Ibid.

by AI beyond a classification of usage patterns and domains.¹⁷ Greater transparency would assist in handling the ethical concerns with using AI for decision making outlined below.

Recommendation 1: Transparency statements include the requirement to disclose the kinds of decisions made using AI.

The *EU AI Act* ranks AI uses according to being of high, limited and minimal risk, while some uses of AI that are considered to be of unacceptable risk are completely prohibited. This includes a number of potential future uses of AI by public sector entities including predicting criminal offences and certain applications of surveillance and biometric identification.¹⁸ The Australian Government's *Safe and Responsible AI in Australia Consultation Interim Response* suggests a risk-based approach as well. However, it purports to only separate AI applications into high and low risk, with differences in the regulation of each. It may be worth considering potential future uses of AI by public sector entities, what category they are likely to fall into and whether any potential future applications are of unacceptable risk and need to be prohibited.

Recommendation 2: That potential future uses of AI by public entities are investigated and any that pose an unacceptable risk are prohibited.

The internal frameworks to control bias, discrimination, transparency, and accountability

In recent years, there has been growing concern that <u>bias and discrimination</u> of AI systems have the potential to perpetuate and reinforce existing societal prejudice.¹⁹ If such AI models are deployed in the public sector, the dangers are likely to be magnified.²⁰ Thus, cautious implementation is of utmost priority – both for the competence and accountability of the public sector.

¹⁷ 'Classification System for AI Use', *digital.gov.au* (Web Page)

<https://www.digital.gov.au/policy/ai/resources/use-classification>.

¹⁸ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L 12/7 (*'EU AI Act'*), art 113(a).

¹⁹ Karen Levy, Kyla Chasalow, and Sarah Riley, 'Algorithms and Decision-Making in the Public Sector' (2021) 17 *Annual Review of Law and Social Sciences* 309-334.

²⁰ Bert-Jaap Koops, 'The Concept of Function Creep' (2021) 13(1) Law Innovation and Technology, 29-56.

Inquiry into the use and governance of artificial intelligence systems by public sector entities Submission 31

While evaluating the performance of the AI model during its development is standard practice, evaluating whether a model continues to perform well on new inputs, once deployed, is less common.

This ongoing monitoring is particularly relevant where AI systems are deployed in a different context to the one that they are deployed.²¹ This phenomenon is known as 'shift'. Most simply, shift can occur as the world changes over time. For example, the demographics of a population may change, or the target municipality of the AI model may change. A related concept 'function creep' warns against extending an AI system beyond its original purpose without proper scrutiny.²² The introduction of one AI system may spur new projects not originally planned or lower the barrier to implementing additional uses of data, tools, funding, and other bureaucratic resources now in place. Consequently, it is all the more important that any implementation faces proper scrutiny and ongoing consideration.

In non-algorithmic contexts, government agencies often complete periodic self-evaluation. A common legal approach, for example, is the use of sunset clauses or expiration dates by which a program must be revisited or reauthorized.²³ With regard to AI usage, these approaches are often less common; however, they appear all the more necessary. These evaluations or audit reports should be publicly-accessible, in the form of an online register.

Recommendation 3: Transparency in AI systems, including a 'right to explanation' for AI based decisions.

Transparency forms the cornerstone of a trustworthy government, and yet it is especially at risk in the face of the inherent complexity of AI systems.²⁴ Transparency allows open scrutiny of decision-making processes and the implicit social norms contained within such decisions – and so, transparency forms the essence of democratic legitimacy.

While AI systems, if poorly developed, have the potential to undermine governmental transparency, we should strive to craft AI models that *increase* decision-making transparency.

²¹ Levy (n 19).

²² Koops (n 20).

²³ Felix Mormann, 'Beyond algorithms: Toward a Normative Theory of Automated Regulation' (2021) 62 *Boston College Law Review* 1.

²⁴ K Passig, 'Fünfzig Jahre Black Box [Fifty years black box]' (2017) *Merkur*. Retrieved from <u>https://www.merkur-zeitschrift.de/2017/11/23/fuenfzig-jahre-black-box/</u>.

There has been a growing effort to develop AI models that can explain how and why they came to a certain conclusion – this is to be encouraged.

In line with the EU's GDPR article 13, we recommend that 'meaningful information about the logic involved' in an AI decision must be provided. This condition will ensure public trust in Australian governmental bodies.

Whether there is an adequate line of sight to the output of AI, and the decisions made through its use

The opacity of AI systems is a significant challenge in the utilisation of AI in the public sector.²⁵ Commonly used machine learning techniques allow the system to learn from patterns in data but are a 'black box' to humans as it is not clear what patterns the machine is recognising and using in producing an outcome. AI systems that are created using these techniques are unlikely to provide adequate line of sight to their output and decisions. They may also result in bias by causing a feedback loop, where trends in data are amplified by the training mechanism. This is particularly concerning in the public sector, where accountability, transparency and impartiality are crucial.

Policies and processes that promote transparency, contestability and oversight have been identified as the best ways to mitigate bias in AI.²⁶ Having an adequate line of sight to the output of AI is crucial to ensure applications of AI in the public sector are appropriate. AI use in government should be explainable including understanding how inputs and variables influence the outcome.²⁷ Contestability is also essential where AI has the potential to significantly impact people or the environment.²⁸ Ensuring avenues for individuals and groups to voice concerns about the impacts of AI use by public sector entities is important.²⁹

Recommendation 4: That legislative protections are implemented for individuals to contest or seek human review of decisions made by Al.

- ²⁸ Ibid 23.
- ²⁹ Ibid.

²⁵ Artificial Intelligence in the Public Sector (n 1).

²⁶ Ibid 34.

²⁷ Australian Government et al, *National Framework for the Assurance of Artificial Intelligence in Government* (2024) 21.

While explainability and options for recourse for individuals who believe an AI system has made a mistake are important, it is also essential to prevent harmful decisions in high risk areas from happening in the first place. Therefore, adequate line of sight to the outputs of AI is crucial before and during its implementation where it is responsible for decision making with minimal human oversight. Transparency about how a decision occurred after the fact is essential but not sufficient. Both preventing harm and providing sufficient responses where harm does occur as a result of AI use is crucial.³⁰

The Australian Government's *Safe and responsible AI in Australia consultation* interim response endorses a risk-based approach to regulating AI.³¹ Such an approach would allow the development and use of low-risk AI with minimal restrictions while AI with higher risk of harm would be subject to different regulations.³² Such an approach may facilitate different line of sight requirements depending on the potential for harm in a given AI system. Appropriate oversight is necessary where potential harm resulting from AI is sufficiently serious or sufficiently likely. This may look like ensuring that 'black box' AI tools are not used to make final decisions on behalf of public sector entities in high risk areas.

Recommendation 5: That the use of 'black box' AI in the public sector is regulated such that it cannot be used without significant human oversight for high risk applications.

Robodebt is an example of the potential for significant harm and the compromising of public service values where there is inadequate line of sight to the output of automated decisions. The 'opaque government system' compared fortnightly income reported to Centrelink with averaged annual figures and if they didn't line up, a welfare debt was automatically generated without human investigation.³³ The *Royal Commission into the Robodebt Scheme* found that the scheme demonstrated the need for oversight in automated decision making and suggested an independent oversight entity be created to review government automated decision making.³⁴

³⁰ Australian Government Department of Industry, Science and Resources, *Safe and Responsible AI in Australia Consultation - Australian Government's Interim Response* (2024) 5.

³¹ Ibid 13.

³² Ibid 13.

³³ Luke Henriques-Gomes, 'Robodebt: Five Years of Lies, Mistakes and Failures That Caused a \$1.8bn Scandal', *The Guardian* (online, 11 March 2023)

<https://www.theguardian.com/australia-news/2023/mar/11/robodebt-five-years-of-lies-mistakes-and-failures-that-c aused-a-18bn-scandal>.

³⁴ Royal Commission into the Robodebt Scheme (Final Report, July 2023).

Recommendation 6: That an independent oversight entity is created to review automated decision making in the public sector.

Whether internal governance structures that currently exist for AI will ensure its ethical and responsible use by public sector entities

The Commonwealth has developed the Voluntary AI Safety Standards (**the Standards**) to guide all Australian organisations in innovating, adopting and using artificial intelligence. The Standards were released on 5 September 2024 and are informed by Australia's eight AI principles. The Standards and principles currently represent Whole of Government assurance in regards to regulating ethical use of AI in the public sectors.

The Standards are voluntary and do not contain any mandatory compliance requirements. For example, transparency is one of the eight principles government agencies are encouraged to consider. However, as the standards are voluntary and exist to encourage ethical conduct, there are no requirements for agencies to report to the Digital Transformation Agency, another public body nor publish for public view AI tools used or developed by an agency.

Governance measures encourage considerations of ethics and responsible use but lack regulatory measures aimed at mitigating and minimising adverse outcomes, particularly in regards to vulnerable members of the public. Currently, there are only AI non-specific laws that vaguely address and regulate certain aspects that AI shares with other subject-matters. These laws and regulations include privacy, competition and consumer protection law, corporations law, online safety, anti-discrimination and human rights, criminal and sector specific laws. Without clear consistent guidelines with mandatory and binding legal effects on agencies, there is risk of adverse outcomes.

Recommendation 7: That the Committee recommend the Parliament create a regulatory body to monitor AI development and use by government agencies, develop AI regulation laws, and investigate unethical practice and tools.