Submission to the Consultation Paper on the use of Automated Decision-Making by Government

January 2025

percapita
FIGHTING INEQUALITY IN AUSTRALIA

The Centre of the Public Square
A Per Capita initiative

SUBMISSION TO THE CONSULTATION PAPER ON THE USE OF AUTOMATED DECISION-MAKING BY GOVERNMENT

The Centre of the Public Square (CPS) at Per Capita thanks the Attorney General's Department and welcomes the opportunity to provide a submission on this consultation.

Per Capita is an independent think tank, dedicated to fighting inequality in Australia. We work to build a new vision for Australia, based on fairness, shared prosperity, and social justice. The Centre of the Public Square works to create equity and fairness for Australians online by holding technology companies to account and building better models of citizen collaboration by imagining new methodologies and alternate technologies for the Australian public.

This submission outlines our recommendations on how we ensure automated decision-making (ADM) in government is fit for the AI age, and that rules and frameworks stay relevant despite rapid technological development. In particular we will address these questions raised in the consultation paper:

- How should the need for transparency about the use of ADM be balanced with the need to protect sensitive information about the business process and systems?
- What transparency rules would be appropriate to build into the framework?
- Should safeguards be different depending on the risks associated with the use of ADM for a particular decision or administrative action?

Summary

Robodebt is one of the most significant stains in public administration in recent history. It is imperative that Australia and the public service learn from this incident and ensure it doesn't happen again.

Even though Robodebt occurred only a few years ago, the technological landscape has changed significantly during that time. Automation is increasingly commonplace and being adopted in the public service, and in society in general. Artificial intelligence has exploded in the last few years, which impacts on algorithmic decision-making and automation more broadly.

The skills, expertise and werewithal required to navigate, let alone fully understand complex AI and ADM systems today is increasingly difficult.

It is in this environment that we have to consider regulation which accounts for the rapid development of ADM systems and algorithms, and ensure any proposals are still up to date and relevant, with the privacy and protections of the Australian public firmly in mind.

We recommend therefore, that:

1) **Regulation around ADM clearly accounts for artificial intelligence (AI)** vs. simple algorithms that don't have learning or iterative capabilities, and have an increased risk threshold for AI activity
2) **That a core tenet of ADM includes interpretability** – the ability for any end user to clearly interpret or explain the processes, logic and eventual result which the algorithms have developed
3) **Core privacy principles are adopted to apply to ADM**, thereby protecting individuals' data privacy and rights as part of ADM systems

Simple algorithms vs. Artificial intelligence (AI)

A key distinction between 'simple' algorithms and AI is our ability to clearly define the inputs and expected outputs of each model.

As a general rule, a 'simple' algorithm is coded by a human programmer, using specific, clearly defined instructions. Its outputs should be clearly identifiable as well once checked against the original coded instructions.

Artificial Intelligence may have been coded by human programmers, but has also subsequently been trained on different models and fed a vast amount of data (often from unknown sources), which facilitate novel developments and novel conclusions. The process and systems whereby AI develop its reasoning, pattern making and eventual results are currently unknowable.[1] Even the owners of the largest AI companies today, openly admit that they do not know how AI comes up the results that it does.[2]

The risk threshold therefore, between 'simple' algorithms and AI are large. Our ability to clearly differentiate between the two are important, but is increasingly getting more challenging. This consultation should be very mindful of the current regulatory process happening for AI, and clearly define its position and relevance as it relates to AI systems.

One of the only ways to clearly differentiate between simple and complex ADM would be to identify the source software. Understanding whether the algorithm being run is a bespoke, government owned program, or an outsourced one from a private, commercial company overseas, or a combination of both would go towards being able to account for the systems and processes in place.

This should be a key requirement as part of the "Publication of general information about the use of ADM" being considered.[3] This should include: source software (the specific software and versions being run) and source developer (who created the software – including any external, non-governmental vendors and suppliers).

---

[1] Snyder, A. (2024). *Shedding light on AI's black box,* Axios, https://www.axios.com/2024/08/22/ai-llms-black-box-interpretability-evaluation

[2] Tangerman, V. (2024), *Sam Altman admits that OpenAI doesn't actually understand how its AI works,* Futurism, https://futurism.com/sam-altman-admits-openai-understand-ai

[3] Australian Government Attorney-General's Department (2024), Use of automated decision-making by government Consultation Paper

Part of the challenge of regulating technology is the perceived complexity and opacity of its systems, with technology often being left to 'tech people' to manage and understand. Even if the systems around it are complex, every piece of software infrastructure at its core is running a particular software and software version, and made/maintained by a particular developer. Understanding the source means we are able to assign accountability for the software, and ideally, that there is a trail of accountability able to be put in place should something go wrong.

Interpretability vs. Transparency

While transparency has been clearly identified and thought through in the Consultation paper, with various sensible recommendations and options put forward, there is an important gap that has been missed.

Interpretability, also sometimes referred to as 'Explainability' is our ability to clearly define and interpret/explain algorithmic models, processes, systems and outputs in a way that make sense to a human.

This has been identified as an issue in the paper, describing how "in some circumstances, it may not be possible to provide full transparency about the specific coding of the ADM system or algorithmic model due to difficulties explaining complex algorithmic processes in a way that is meaningful to general members of the public".[4] However it has not been addressed.

When it comes to AI systems, this gap is even more problematic. Even the CEOs and founders of the largest AI companies admit that they do not know how their AI models run and come up with the results it comes up with. Most AI systems running today are black boxes and its inner workings are not transparent or interpretable.[5]

Even if this consultation makes a distinction between 'simple' algorithms and more complex AI, the general members of the public will likely still have trouble understanding the technical methodologies and techniques involved.

Transparency therefore, would be meaningless without an ability to explain and understand the ADM systems and processes being revealed. Interpretability must go hand in hand with any transparency initiatives, so that there are accessible descriptors, labels, sample case studies, explanations, intended use cases, user personas, logic flows and scenarios accompanying any ADM systems. Ideally, there should be case studies that demonstrate intended outcomes and results, and unintended outcomes and results, so that there is an overall spectrum made clear to determine if the ADM systems are functioning as intended.

There should also be accessible training and education sessions more broadly, free to use for the Australian public, which provides more information and clarity on ADM systems being used as part of government and government services.

---

[4] Australian Government Attorney-General's Department (2024), Use of automated decision-making by government Consultation Paper

[5] Daniel, W. (2024), *Google CEO Sundar Pichai says 'hallucination problems' still plague AI tech and he doesn't know why*, Fortune, https://fortune.com/2023/04/17/google-ceo-sundar-pichai-artificial-intelligence-bard-hallucinations-unsolved/

Core data privacy principles

A fundamental way to minimise risk and prevent ADM errors would be to enforce data privacy rights so that there are privacy protections in place that would apply to data capture and algorithms.

There is currently a wholesale review of the Privacy Act in progress, and this consultation should apply relevant principles that would impact.

Principles like data minimisation, purpose limitations and penalties for breaches would create important protections for individuals in ADM systems.

Data minimisation is an approach which seeks to minimise the amount of necessary data being captured about individuals, capturing data only as it relates to the service or function that it's being collected for.

This is related to the idea of purpose limitations which is about using that specific data only in the specific context and use case that it was intended for, and nothing else.

Further, penalties around breaches would discourage any liable persons from wrongdoing. The final findings of the Robodebt Inquiry found that there were 12 public servants in breach of codes of conduct.[6] However, there were no consequences for those 12 public servants, and none of them had their roles terminated.[7]

For such a consequential program, which harmed thousands of Australians, there must be clear penalties involved. Currently the Privacy Act Review is considering a tort for serious breaches, and any misuse of ADM systems should strongly consider enforcement mechanisms and penalties as well.

By upholding data privacy rights and applying principles like data minimisation and purpose limitations, as well as penalties for breaches, ADM systems will have built-in standards to minimise the risks of unintended consequences like Robodebt.

[6] Australian Public Service Commission (2024), *Statement by the Australian Public Service Commissioner on the Robodebt Centralised Code of Conduct Inquiry*, https://www.apsc.gov.au/about-us/working-commission/who-we-are/media-releases-and-statements/statement-australian-public-service-commissioner-robodebt-centralised-code-conduct-inquiry

[7] Bajowski, J. (2024), *'It's not an insider game': APSC confirms no public servants sacked over robodebt*, The Mandarin, https://www.themandarin.com.au/254767-its-not-an-insider-game-apsc-confirms-no-public-servants-sacked-over-robodebt/