HYPEREAL AtomEthics

Submission to Attorney General’s consultation questions on the use of automated decision-making in government

15 January 2025

About Us

Hypereal helps governments of all sizes to design and implement simple, clear and fast digital services and platforms that benefit citizens and industry, and to think through the opportunities and challenges of operating in the digital age.

Hypereal has particular expertise in data and AI ethics, including helping to shape the widely adopted NSW AI Assurance framework, and delivering Services Australia’s post-Robodebt Data Trust & Ethics framework. Our AtomEthics platform supports ADM and AI service builders and owners to address design and risk themes, including those expressed here.

Introduction

The consultation paper to which this submission responds addresses possibilities for legislative harmonisation in the context of automated decision-making (ADM) within government. The purpose of the paper is to assist in formulating responses to the two recommendations of the Robodebt Royal Commission that focus on ADM, and its intent to consider all forms of machine-based decision support and decision-making is to be applauded. We are hopeful that the consultation’s outcomes will contribute to setting clearer boundaries for the ethical design and use of process and decision automation in government, including in new use-cases that have yet to surface.

However, Robodebt represents a complex and still evolving sociotechnical failure that better laws and clearer technology parameters alone would not have prevented. It is important to acknowledge that regulatory and legislative reform will not offer complete protection against the harms of schemes like Robodebt and similar global counterparts. In isolation, these would not have prevented the design and implementation of a government service that was intentionally malign (Thompson et al, 2024). The combinations of organisational culture and underpinning belief structures that can give rise to deliberatively harmful organisational strategies must also be better understood, monitored, and addressed.

A stronger safeguarding system will incorporate these elements, and particularly valorise the moral intuitions and legislative knowledge of the frontline workers who are most proximate to an operating ADM system. One of the risks outlined in the consultation paper is the de-skilling of human operatives and subsequent over-reliance on ADM. It should not be overlooked that in the case of Robodebt, it was these very operatives who spoke most cogently against it and who must form part of an organisation's safety culture. Relatedly, the risks highlighted in the consultation paper are narrowly focussed on bias, coding errors, and inaccuracies in rules-as-code. We invite you to extend your gaze to the many other elements of digital service design that feed into machine-based decision support and determinations (some of which are outlined in our response to Q9 below) as critical elements that your reforms could capture.

We thank you for the opportunity to submit and would be pleased to engage further.

1. **How should the need for transparency about the use of ADM be balanced with the need to protect sensitive information about business processes and systems?**

   The need to protect sensitive information that relates to systems and business processes is not well articulated in the consultation paper. We suggest that, like other jurisdictions (e.g. https://interoperable-europe.ec.europa.eu/collection/open-source-observatory-osor/news/uk-government-prescribes-open-source-public-procurement) government should open source services by default in order to promote robust, reuseable government software that saves money and is defensible both in and out of court. Open sourcing government services is one of the underpinning principles of open government, the overall policy aims of which are to improve both government accountability and public participation.

   This principle of transparency is known in the technology space as "working in the open". Its purpose is to show that the workings of government are trustworthy. In addition to code, transparency also requires that information is fully available in formats that people understand. It is, for example, laudable that the consultation paper is available in a "plain language" version: however, this 2-page version omits significant detail contained in the longer version on which we base our response. Transparency assists people to understand the processes and deliberations by which government digital services arrive at their determinations. It also helps people understand how laws are articulated as processes and practices, including for the purposes of testing their interpretation.

The philosophy common to open government and the transparency principle is that they will provide for debate that strengthens policy and service outcomes. Adopting the transparency principle does not simply mean providing information but also that mechanisms are developed for capturing and contextualising feedback and reflecting it back as service and policy amendments.

Robodebt, to which this consultation paper is a response, highlights what can happen when transparency is not privileged either in service design or in legal processes. Recipients of robodebts received no explanation of the reason for or composition of their debt, and many were required to submit FoI requests to obtain their files, which on receipt contained information that was challenging to interpret. In the context of legal processes and parliamentary inquiry, transparency was deliberately avoided in order to prolong the operation of the Scheme or to obscure accountability for it. Debts for the two test cases (excepting the ATO interest in *Amato*) were zeroed out on "review" in order to claim that there was no longer a case to be made and avoiding the documentary disclosures that would have followed. Adverse rulings at AAT1 were not contested because this would have meant an open-court AAT2 hearing and the possibility of journalists in attendance. The class action was settled on the day proceedings were to commence, meaning that key documents relating to legality and knowledge remained out of public view, coming to light only during the Robodebt Royal Commission. The second Senate Enquiry was repeatedly rebuffed in its quest for those documents on the pretexts of public interest and cabinet confidentiality. Only Robodebt Royal Commission's powers of compulsion enabled the truth to become known, and, as a downstream consequence, for this present consultation to occur.

2. What transparency rules would be appropriate to build into the framework?

The framework will need to accommodate two levels of transparency: 1) the recipient of the ADM enabled outcome and 2) monitoring and audit oversight functions. A detailed scope of transparency rules is outside the parameters of this preliminary response. However, they should include:

- easy access to, and interpretation of, a flow or other representation of the process and the points at which ADM is applied or is assisting human decision-making;
- the laws/ regulations or other in which the process is grounded;
- the data sources on which the decision is based;
- the parameters informing the decision;
- how those parameters were combined, weighted, or assessed, and
- the appeals process, including an easy prompt to lodge and appeal.

Monitoring and auditing will require further parameters that will depend on the nature of the service but is likely to include such things as:

- the authorising entity for the ADM system;
- the conditions under which the system will be rolled back and by whose authority
- the conditions under which the system should be fully reevaluated (e.g. a change from an advisory to a compliance model);
- datasets, sources, lineage, treatment, and combination
- model construction;
- confidence intervals;
- model precision, recall, performance and rebalancing history;
- the overall volume of decisions handled by the system and their distribution, and
- system feedback in the form of appeals, both successful and unsuccessful
- data elements that are collected but not used in the formulation of a decision

Many of these parameters should also be publicly reported as performance indicators for the digital service.

3. What pre-implementation safeguards should apply where ADM is intended to be used?

This is a difficult question for us to answer succinctly. We have built a very complete digital data and AI ethics governance platform that considers, at each point in its lifecycle from initiation to decommissioning, the safeguarding perspectives with which designers and system owners need to have engaged. This lifecycle approach, consistently with the recommendations of Ng & Gray (2022) cited in the consultation paper, provides for a a more nuanced and dynamic means of safeguarding than a one-off pre-implementation assessment.

Pre-implementation encompasses more than one lifecycle phase. Safeguarding factors are phase-dependent and often cumulative. At the initiation phase, for example, safeguarding considerations are those that establish foundational good practices and identify gaps and factors that elevate other safeguarding risks. Skipping ahead to the later data phase, safeguards should focus on the nature of consent, the data itself, its sources and treatment, and its proportionate use.

We also advocate for a gap-based approach over the more common risk matrix approach. This means that safeguarding regimes should consider the posture and response implied by formal policy positions, coverage gaps in policy relative to the selected safeguards, and the implications of upcoming changes to the external environment, including legislation. Lastly, safeguarding regimes must include processes for monitoring and auditing. We are happy to engage further on detailed parameters for consideration should you deem it appropriate.

4. What system-level safeguards should be required, to ensure that ADM operates appropriately?

See our responses to Q2 and 3 for proposals.

The consultation paper contemplates the establishment of a framework that is then (possibly) followed by the establishment of an expert oversight body. In our opinion, the establishment of an expert body should precede the framework, because the framework will be made more robust and implementable by the inclusion of multi-disciplinary perspectives and practical expertise of those who have direct experience of designing and managing safeguards in high-risk, high-consequence settings such as the provision of programmes of social protection.

Additional non-technology based safeguards must be considered (per other Royal Commission recommendations) for better mitigation of ADM risk, including increasing the independence of formal and informal regulators at all levels, and as specified in our Introduction.

5. **What decision-level safeguards should there be for persons affected by decisions made using ADM (for example, review rights)?**

The right to appeal is a foundational right for all people who are subject to an adverse decision, including one reached by ADM. Review processes also provide feedback and a healthy corrective into a system that can be used to improve its functioning.

It is also appropriate to institute a sampling regime for a review of randomly selected recommendations and decisions by a human. This precaution (together with the reporting parameters outlined in Q2) will help to detect where a system is generating inconsistent or incorrect advice or outcomes.

6. **Should individuals be notified of the use of ADM? Yes/ No Please expand on your response**

Yes. Individuals have a right to know whether their outcomes are algorithmically determined. This right must be combined with the transparency parameters expressed in response to Q2.

The notification should be at the point immediately prior to engaging with the system. General wording on a departmental website will not be timely nor will it connect appropriately with the person using the service.

7. **If yes, should notification be required at a specific point in the decision-making process, or should flexibility be provided to agencies about the appropriate time to make a notification? Yes/No Please expand on your response**

No. The question is a little ambiguously worded, but no, there should be no flexibility for agencies to determine when in the user journey to flag the use of ADM. This must be at the beginning of the user's engagement with the system.

The report of the Robodebt Royal Commission highlighted how many of its victims were now too frightened of engagement with government systems to obtain services which they and their families needed, and to which they were entitled. Government therefore needs to foster design patterns that will become familiar and reassuring to citizens across the range of services they access. These patterns should include early and full disclosure of how and why ADM is used in a service before the citizen engages with it. The intention of this practice is to increase a person's confidence in the system.

This notification should be re-confirmed when the decision is delivered and should be accompanied by an explanation and by a prompt to start the appeal process as necessary.

8. **Should there be any exemptions to ADM safeguards? If yes, what exemptions should be included and why? Yes/No Please expand on your response**

   No (with caveat). Any ADM safeguard exemptions should relate uniquely to issues of national security (and the incidence of such exceptions should be reported). Safeguards are established to protect the vulnerable. Exemptions to their application therefore represent a weakening of these protections.

9. **Should the safeguards be different depending on the risks associated with the use of ADM for a particular decision or administrative action? Yes/No Please expand on your response**

   No. Applying the same safeguarding level to all services ensures robust consistency within the safeguarding process. This guarantees that, when a decision is taken that a specific service does not require the highest level of safeguarding, that is made actively and is documented. This process need not be a lengthy or bureaucratic, but it does mean that an auditable system of record is kept for each ADM instance.

   Robodebt highlighted how the parameters of automation can create sometimes fatal risks for vulnerable citizens, including limiting vulnerability flags, excluding relevant documents that have already been presented, requiring evidence that may not be available, imposing sanctions and fines, creating time pressure, and threatening legal action. It is important to recognise that each of these represented a design choice and the potential for a unique or cumulative failure in a safeguarding regime that is not acknowledged within your current risk set overview.

   Additionally, it should be recognised that human-in-the-loop is a key component of safeguarding, particularly in high-risk, high-consequence settings such as programmes of social protection.

References

Ng, Y. F., & Gray, S. (2022). Disadvantage and the automated decision. Adel. L. Rev., 43, 641.

Thompson, C., Samson, D., & Kurnia, S. (2024). Neither fair nor legal. How and why untrustworthy digital ecosystems evolve. Scandinavian Journal of Information Systems, 36(2), 8.

Enquiries

hello@hyperealhq.com
0417 131 800