3 February 2022
Royal Commission into the Robodebt Scheme

By Submission Form

Submission - Royal Commission into the Robodebt Scheme

We are grateful for the opportunity to provide a public submission to the Royal Commission (**Commission**) into the Robodebt scheme. This is a joint submission prepared by the Australasian Society for Computers and Law (**AUSCL**), the UNSW Allens Hub for Technology, Law and Innovation (**Allens Hub**), and independent subject matter expert contributors, and endorsed by the Gradient Institute.

The submission provides an expert view from digital transformation, automation, legal technology, public sector and other relevant experts to the Commission on the role of automation and automated decision-making in government and, in particular, how these technologies can be used safely and responsibly to deliver good outcomes for Australian communities. It is also worth noting that the recommendations could also act as a blueprint for other countries, many of whom are already exploring the use of Artificial Intelligence in social protection programmes, similar welfare systems, and with comparable data infrastructure.[1]

In the submission, we set out key criteria for establishing trustworthy automated systems in government contexts and discuss how, if these criteria had been applied to the Robodebt scheme, the harms created by the scheme could have been avoided.

About us

AUSCL is an interdisciplinary network of professionals and academics focussed on issues arising at the intersection of technology, law and society. It is a registered Australian non-profit charity with a charter to advance education and advocacy. AUSCL was officially launched in July 2020, but its member State societies were formed as early as 1981. AUSCL provides a

---

[1] Lokshin, M., & Umapathi, N. (2022). AI for social protection: Mind the people. Brookings Institute

forum for learned discussion and debate through its Policy Lab, Working Groups and Events Program attracting support and engagement across Australia and globally.

The Allens Hub for Technology, Law and Innovation is an independent community of scholars based at UNSW Sydney. As a partnership between Allens and UNSW Law, the Allens Hub aims to add depth to research on the diverse interactions among technology, law, and society. The partnership enriches academic and policy debates and drives considered reform of law and practice through engagement with the legal profession, the judiciary, government, industry, civil society and the broader community. More information about the Allens Hub can be found at http://www.allenshub.unsw.edu.au/. The submission represents the views of the authors and is not an institutional position of Allens or UNSW.

The Gradient Institute is an independent, nonprofit research institute that works to build ethics, accountability and transparency into AI systems: developing new algorithms, training organisations operating AI systems and providing technical guidance for AI policy development.

Introduction

The Robodebt scheme has exposed significant cultural, structural, and political issues in how public institutions have administered and managed public policy. These issues were compounded by how technology has been misused to deliver public harm at scale.

The goal of this submission is to support meaningful and pragmatic change in the Australian Public Service (and by extension, all public sectors) with recommendations about the issues observed in the Robodebt scheme relating to a) ***poorly designed and poorly implemented automation***, b) the dangerous and systemic ***inability to understand the human impact of government policies*** and take appropriate action to avoid and mitigate harm, and c) ***the need for an adaptive and continuously evolving approach to policy*** informed by feedback loops, evidence and a culture of service to the public and stewardship of sustainable public good.

Observations

The Robodebt scheme was not an example of the dangers of automation, but of poorly designed and implemented automation.

Automated systems are often deployed to facilitate cheaper delivery, rather than better outcomes for users. However, it is possible and practical to design automated systems to deliver more efficient and cheaper outcomes while still delivering benefits to the community including faster decisions and increased transparency. Further, this can (and should) be done in line with established principles of administrative law.

To achieve this, we propose that automated or automation-enabled processes in government should be required to be designed to satisfy certain key design criteria. The Rules as Code

(**RaC**) movement[2] can offer some guidance on how to build trustworthy automated decision-making processes. RaC practitioners suggest the following key requirements:

1. **Transparent** - the rules governing the decision should be visible to the public, ideally in an accessible format, and able to be interrogated.

2. **Traceable** - It should be clear what specific rules (legislation, regulation, case law, etc) were applied, and what facts and evidence those rules were applied to. This provides a foundation for explaining decisions.

3. **Accountable** - it should be clear and documented which entity has made a decision (e.g., department or agency), and how the decision is within the legal mandate of the entity. The burden of proof should clearly lie with the decision-making entity. The conditions under which an automated decision is taken to be valid can also support other requirements (such as transparency).

4. **Appealable** - the subject of the decision should be able to seek a review of the decision (for example, regarding an error of fact, or if they believe a rule has been incorrectly applied).

5. **Beneficial** - decisions and decision-making systems should prioritise public benefit, beyond mere cost savings or efficiencies; see Recommendation b) below).

Further, administrative law has its own requirements which can be implemented as technical requirements, including but not limited to:

1. The placing of the burden of proof on the administrative decision-maker, not on the subjects of the decision.

2. The requirement to only take into account relevant evidence, and not consider irrelevant matters, and avoid bias.

3. The requirement to provide reasons for decisions (which can be enabled by traceable decision-making), and procedural fairness and accountability mechanisms (which can be enabled by appeals functionality).

These and other requirements for automated systems can be set out in authorising legislation.

Additionally, it is critical that the drive towards increased automation does not displace the exercise of discretion. Automated systems are most appropriate for dealing with prescriptive rules and, in most government decisions, there is a need for the exercise of discretion to enable procedural fairness. We observe that in the Robodebt context, there was no effective access to human arbiters, as there was no functional appeals process. The goal should be to enable *better* decisions, not just faster and cheaper decisions - in some cases partial automation with the final decision made by a human arbiter is the ideal outcome.

Further detail and references are set out in the Appendix.

---

[2] For background on Rules as Code, see, e.g., Mohun, J. and A. Roberts (2020), "Cracking the code: Rulemaking for humans and machines", *OECD Working Papers on Public Governance*, No. 42, OECD Publishing, Paris, https://doi.org/10.1787/3afe6ba5-en.

The inability to understand human impacts of government policy leads all policies to be potentially harmful

The financial *and* human impacts of policy and implementing systems should be identified, measured and continuously monitored, alongside other performance measures. In this way, governments can identify whether policy intent has been effectively achieved and detect and mitigate harm if and when it arises. There is local and international precedent for this - see, for example, the [NSW Government Human Services Outcomes Framework](#). Measurement and monitoring for human impact needs to also be complemented by a more holistic and adaptive approach to managing public policy delivery, including escalation of issues (such as harm) as they are identified, and proactive iteration or mitigation of policies to improve human impacts.

The safe and responsible uses of new technologies in government requires a culture receptive to feedback and that prioritises continuous improvement

The Commission has heard extensive evidence that multiple attempts to alert decision makers in the Department of Human Services (**Department**) to concerns over the legalities and impacts of the Robodebt scheme, by both internal staff and external advisors, were not met with a receptive audience. There are many potential reasons for this, but it is clear that a Ministerial/Government objective was continuously prioritised over the Departmental purpose, mandate, legislative constraints and duty of care, from the top of the Department down.

We now live in a digital world, but the basic requirements of "responsible government" have not changed. People expect transparency, accountability and public engagement in order to have confidence in and engage with the important work of the public sector. To successfully implement policy through technology, it is critical to also develop a culture that proactively identifies flaws and bugs, and which actively seeks feedback and enables escalation to manage risks and implement mitigations. It would be helpful for the SES culture to learn from modern digital practices of continuous iteration and improvement, and to learn and be accountable to the principles and requirements of Administrative Law for all that they administer.

Policy should be evidenced-based, and policy makers should be capable of pivoting and evolving policy and implementation when new evidence is discovered. Policy owners must also be capable of alerting Ministers of policies that are illegal, potentially harmful or inappropriate, and being confident to uphold their statutory responsibilities, so as to grow and maintain public confidence in the public sector. After all, such confidence is critical for perceived legitimacy of all that the sector administers, from policies to services, and even elections.

Recommendations

Our recommendations (with supporting information in the Appendix) are as follows.

a) to address ***poorly designed and poorly implemented automation*** by doing the following:

- Establish key design criteria to support trustworthy and lawful decision-making, compliant with the principles of administrative law and rule of law for all automation

systems in government, especially those that substitute for human decisions. This could include, for example, establishing a legal "right to an explanation" similar to [Section 23 of the New Zealand Official Information Act,](#) which would create a legislative requirement to design explainable systems.

- Establish test-driven approaches to policy change, including the Better Rules methodology where Rules as Code are drafted simultaneously with legislation for modelling, testing and assurance.

- Establish practical guidelines and assurance mechanisms for automation systems in government, including a public catalogue of departmental use of automation, AI and algorithms, with publicly available information about risk and impacts. The [Canadian Algorithmic Impact Assessment Framework](#) could provide a useful starting point. Relevant work is also being commissioned [by the NSW Ombudsman for that state](#).

- Ensure all appeals and feedback are monitored holistically for patterns of issues, including independent oversight from outside the Department.

- Build into the [recently announced "Charter of Partnerships and Engagement"](#) a requirement for departments to engage on policy and automation system design, ideally with those directly affected by such systems or representative bodies. This should extend to engagement on proposed oversight, governance and appeal mechanisms and evaluations thereof.

b) to address the dangerous and systemic ***inability to understand the human impact of government policies*** and take appropriate action to avoid and mitigate harm:

- Develop and embed a methodology for identifying and measuring human impacts. We note that work is underway on a new international standard for impact assessment for AI systems (Australia is participating in development work through Standards Australia).

- Establish a whole of government function where all policy interventions are monitored, both for intended policy impacts, and for human impacts. Holistic impact monitoring enables analysis of unintended consequences from change. Such a function could advise on harmful patterns and would oversee policy reforms to address harms.[3]

- Require the current Regulatory Impact Statement requirement for new policy proposals be changed to "Impact Statement", and to explicitly include human and community impacts, not just economic or regulatory ones.

- Establish a publicly available digital representation of social and taxation policies (a "Policy Twin"), to support public testing of system outputs against the law, and to democratise policy modelling and monitoring.

c) to address ***the need for an adaptive and continuously evolving approach to policy***, informed by feedback loops, evidence and a culture of service to the public and stewardship of sustainable public good:

---

[3] For further information about the opportunities and challenges of policy monitoring, see Schoenefeld, Jonas J., et al (2019), "Policy Monitoring in the EU: The Impact of Institutions, Implementation, and Quality", https://link.springer.com/article/10.1007/s11615-019-00209-2

- Establish mandatory training of all public servants and those working with the sector on the responsibilities, accountabilities and special context of the public sector, including to deliver public value, accountable stewardship, and the practical implications and considerations of Responsible Government on all aspects of the sector.[4]

- Establish a safe and optionally anonymous feedback tool on policy and interventions, to close the knowledge gap between policy decision and delivery, and provide more accountability to understand and address the impacts of government policies. This would enable feedback from public servants without fear of reprisal, enabling early intervention.

- Update the Australian Policy Handbook to include end-to-end policy lifecycle management in the age of digitally-enabled policy delivery, with multi-disciplinary policy teams, test driven approaches to policy interventions, real time monitoring, and feedback loops to continuously improve policies based on real world impacts and evidence.

Primary Authors

Tim de Sousa, CIPM, CIPP/E, FIP - AusCL Future Law Workstream

Pia Andrews - Public sector transformation expert (in a personal professional capacity)

Lyria Bennet Moses - Director, the Allens Hub for Technology, Law and Innovation

Endorsements

The Gradient Institute

Appendix - Applying the Key Criteria for Trustworthy Automated Decisions to the Australian Federal Government's 'Online Compliance Intervention' program ('Robodebt')

The Key Criteria for trustworthy automated decisions in government

Decisions should be:

1. **Transparent** – the rules governing the decision should be visible to the public, ideally in an accessible format, and able to be interrogated.

2. **Traceable** – It should be clear what specific rules (legislation, regulation, case law, etc) were applied, and what facts and evidence those rules were applied to. The steps in the decision-making process should be explainable and auditable, and clear to the subject of the decision.

3. **Accountable** – it should be clear and documented which entity has made a decision (e.g., department or agency), and how the decision is within the legal mandate of the entity. The conditions under which an automated decision is taken to be valid can also support other requirements (such as transparency).

4. **Appealable** - the subject of the decision should be able to seek a review of the decision (for example, regarding an error of fact, or if they believe a rule has been incorrectly applied).

5. **Beneficial** - In addition to the above, decisions and decision-making systems should prioritise public benefit, beyond mere cost savings or efficiencies. This means the financial *and* human impacts of policy and systems should be identified, measured and monitored, alongside other performance measures (including the avoidance or mitigation of harm), to identify whether policy intent has been effectively achieved and to enable the detection and mitigation of harm if and when it arises.

Read more:

- [*Rules as Code Handbook: Building Trustworthy Systems Using Coded Rules*](#)

The purpose of the Key Criteria

Given the potential cost savings and efficiency benefits of automating government decision-making, it is highly likely that governments will be motivated to implement more automated decisioning processes.

As such, we must ensure that governments are building decision-making systems that mimic the properties of the most robust human-made decisions. In government, this is a simpler proposition to justify than in a business environment. Specifically, the principles of administrative law provide a clear framework around which to build decision-making systems. That is, machine-enabled decisions should be explainable and be explained, all relevant matters should be considered, irrelevant matters should not be considered, and decisions should be based on facts for which there is evidence. If building systems to make administrative decisions, governments are already required to ensure that those systems comply with the principles of administrative law.

It is possible to build decision-making systems that comply with administrative law principles – this is the purpose of the Key Criteria. When implemented correctly, an automated decision should be at least as robust as human-made decisions, if not more robust. For example, it should be technically possible to inspect an automated decision-making system to determine (with at least a reasonable degree of confidence) that it has not taken account of any irrelevant matters.

Read more:

- *[Rules as Code Handbook: Why Code Rules?](#)*

- *[A Trust Framework for Government Use of Artificial Intelligence and Automated Decision Making](#)* - Pia Andrews et al, (2022)

- *[Accountability in the age of artificial intelligence: A right to reasons](#)* 94 ALJ 829, Lyria Bennett Moses and Edward Santow (2020)

How could adopting the Key Criteria prevent another Robodebt scenario?

The income-averaging methodology for identifying alleged debts used in the Robodebt program has now clearly been demonstrated as flawed and had a massive error rate from the beginning. That fact aside, had the program adhered to the Key Criteria which we suggest, we consider that the harms of the program could have been largely mitigated.

We have set out our argument below.

1.   Transparent

When we, as a society, enact laws, those laws are [published and made publicly accessible](#).

If we have coded government rules and built a system that uses them, those rules should also be publicly accessible (and, ideally, open source). The code should be inspectable and testable. That way, the community (including independent assessors) can determine for themselves what the rule is and whether the code correctly reflects the rule, and that the system is working correctly and as intended.

Had this been done in the Robodebt scheme, it would have been apparent to and demonstrable by independent assessors from the start of the program that the income-averaging methodology was flawed. Ideally, this would have been established in testing and corrected by program

administrators before the program was implemented, but it clearly was not – accessible rules would have enabled independent assessors to conclusively prove the issue.

2.   Traceable

As discussed above, it is a foundational principle of administrative decision-making that, when making a decision, the decision-maker must give reasons for the decision.

Using systems that use coded rules to make a decision gives us the opportunity to easily and accurately detail how and why decisions have been made. We can design systems to deliver:

- the decision,

- the rules applying to that decision,

- the facts to which the rules were applied, and

- the evidence for those facts on which the system relied.

If this had been done in the Robodebt scenario, when an alleged debt was raised, the Department could have sent out a letter stating:

- the belief that the subject owed a debt – this was stated in Robodebt letters, without further justification

- the rules that were applied in calculating the debt – e.g., "we averaged the income you received in [A] period [in this way]. We have assumed that your monthly averaged income was consistent during [B] period to determine what we believe to be your income during [B] period"

- the evidence that the rules were applied, e.g.:

    o   "ATO data shows that you were paid [A] dollars in income during [B] period";

    o   "Our records show that you received [C] dollars in Centrelink payments during [B] period".

    o   "You were entitled to [E] dollars in Centrelink payments, subject to your income in [B] period".

    o   "If our assumptions about your income during [B] period are correct, you have been overpaid by [F] dollars".

This approach would have enabled the subjects of Robodebt decisions to assure themselves that the right rules and evidence were applied and, therefore, that the decision was correct. It would also have made evident the flawed assumption inherent in income averaging, allowing the person receiving the letter to understand what they needed to correct.

Further, this level of traceability would have better enabled the subjects to appeal the decision (assuming there had been a functional appeals process) and also would have made it easier for the department to determine the validity of the appeal - ultimately this could have significantly reduced the human impacts of the errors of the program.

It's critical that we avoid 'black box' automated decisions in government – where the reasons for the outcome cannot be explained. This includes reliance on machine learning algorithms for which outputs lack explanations as well as rules-based systems subject to commercial-in-confidence restrictions that prevent full transparency as to the rules applied. These 'black box' systems are highly likely to be unsuitable for government decisions as they cannot deliver sufficiently explainable decisions.

Traceable systems with coded, interrogable rules also enable test-driven approaches to policy change - systems can be built to conduct modelling around different variations of rules, testing can be conducted with all possible variables to confirm policy intent is achieved by the application of the rules, and assurance can more easily be conducted once the system is in operation .

Read more:

- '*Better Rules'* approach: *Legislation as Code for New Zealand: Opportunities, Risks and Recommendations* Brainbox Institute - Tom Barraclough, Hamish Fraser, Curtis Barnes (March 2021)

3. Accountable

A decision made by an automated process should be able to be relied upon by the subject of that decision. If a government department outsourced some of its functions to a supplier, the department would remain accountable for those functions. The same should be true of a department which outsources its functions to a machine.

There is a lack of jurisprudence on this issue at the moment, and some counter intuitive outcomes. For example, in *Pintarich v Deputy Commissioner of Taxation* [2018] FCAFC 79, the Full Federal Court found that the automated decision made by the ATO, provided in a letter purportedly signed by the Deputy Commissioner, was not a valid decision. The Court based their finding on the view that there was no 'related mental process involved' in the relevant component of the decision. The outcome was that the ATO was not accountable for the decision set out in the letter provided to Pintarich, and Pintarich could not rely on it, even though Pintarich could not have known that the decision was automated.

The Robodebt letters sent to the subjects of the program did not clearly state that a decision had been made that an alleged debt was owed. Rather, those letters presented the income-averaged debt calculation as an assumed fact without explanation, that had to be disproven. This inappropriately placed the burden of proof on subjects, contrary to principles of administrative laws. The letters did not state the powers under which documents were being required to be produced (and the Government claimed that there was no intent to compel the subjects to provide documentation). The letters also did not set out under which powers the decision was being made, or that the decision (the calculation of the alleged debt) or the debt recovery process was automated. This lack of accountability made it harder for the subjects of the program to understand how to contest the alleged debts.

Read more:

- Janina Boughey and Katie Miller (eds), *The Automated State: Implications, Challenges and Opportunities for Public Law* (Federation Press, 2021).

4.   Appealable

The Department failed to set out a clear appeals mechanism. The Department did not initially permit subjects to speak to office-based Centrelink staff to contest debts, instead requiring that appeals be lodged online via a portal that was frequently not operational. The difficulty of [contacting Centrelink by phone has been well documented](#), and Robodebt subjects and Centrelink staff reported that staff were not trained to handle Robodebt appeals. Further, many recipients first heard of their alleged debt through a 3rd party debt collector, who was not responsible to (or able to) provide any explanation of how the debt was raised, and could not accept appeals.

Had the decisions been fully explainable and traceable, appeals would have been easier to assess and resolve, for both subjects and the Department.

Similarly, all appeals should have been handled in a single system such that patterns of appeals could have been systematically identified and managed. This would have better enabled early identification of program issues.

5.   Beneficial

The policy intent of Robodebt, as reported to the Royal Commission, was to recoup $1.2 billion to the Federal Government Budget. This objective was cited many times by senior Department executives, with Department officials reportedly "particularly unhappy" if the $1.2 billion objective could not be achieved, as the "number itself was a goal of the process".

There is a good argument that the proper policy objective should have been centred around the objectives of social policies, which is about supporting people at their most vulnerable to live well and contribute to their community. Addressing any genuine system inaccuracies in the provision of payments, whether welfare or taxation, needs to be done in the context of public benefit and in a high integrity way that maintains public trust.

Had the Robodebt scheme been designed with the above requirements in mind, then it is possible that an overpayments recovery scheme could have been achieved at least in a more accurate manner, with appropriate and effective access to justice for those subject to debts raised in error. This alone would have avoided or mitigated much of the harm that was reported to the Commission.

Beneficial in the context of government also relates strongly to the confidence the community has in government. Such confidence comes from a willingness to believe that a person or entity is operating in **good faith**, with **integrity**, and in a way that fulfils the individual's **expectations** of that person or entity. Government systems could therefore be considered trustworthy through demonstrating good faith (through a systemic and measurable commitment to human-centred and humane outcomes), by assuring high integrity systems (that are lawful, accurate, high veracity, assured, consistently applied and appealable), and that meets public expectations (by reflecting public values and needs, doing no harm, being transparent and operating within relevant legal, social, moral and jurisdictional limitations of power). Arguably, the Robodebt scheme failed all three of these tests.

Read more:

- *[A Trust Framework for Government Use of Artificial Intelligence and Automated Decision Making](#)* - Pia Andrews et al, Aug 2022.

What next?

Governance

It is critical that governments put in place robust measures to ensure that automated decisions are correct (in the sense that they are based on legal requirements for those decisions) and, moreover, are made in a way deserving of community confidence. Adopting the key criteria for automated decision-making would go some way to ensuring this.

In some other jurisdictions, use of wholly automated decision-making processes has been explicitly regulated - see, for example:

- [Article 22](#) of the EU General Data Protection Regulation prohibits 'wholly automated decisions', except in limited circumstances (for example, such as where the individual explicitly consents). The GDPR requires that measures be implemented to safeguard the rights, freedoms and legitimate interests of the individual. This includes but is not limited to the right to appeal an automated decision to a human arbiter.

- New Zealand has [implemented an algorithm charter](#) which includes similar principles to those set out above, plus additional governance measures such as formalised risk assessments. This charter has been signed up to by most departments, and more measurable mechanisms are being explored to complement it.

- The Government of Canada has established a set of [Digital Standards](#), to guide the development of digital government products and services. Standard 9 is [Design ethical services](#) - this requires that systems and services be designed to ensure that everyone receives fair treatment, and that there is compliance with ethical guidelines in the design and use of systems which automate decision-making. The Standard includes a [mandatory algorithmic impact assessment](#) for any automated decision system. They have also established a [Digital Charter](#), which sets out key rights of Canadians in digital spaces - including openness, consent and control, and ethical use of data. These rights will shape the Canadian Government's future digital products and services, including automated decision-making systems.

In our view, Australia should consider similar requirements, not only for fully automated decision-making (as in the EU) but for all government decisions. Some requirements are only relevant to automation - for example, legislative authorisations that deem the decision of an automated system to be the decision of a Minister or official should specify clear requirements as to the properties of such systems, based on the framework articulated in this submission and in comparable jurisdictions. However, other requirements such as fair treatment, openness, and ethical use of data should apply independent of the extent to which the decision-making process is automated.

Recognition of the limitations of the technology

There needs to be recognition and understanding that automated systems are most appropriate for dealing with prescriptive rules and, in most government decisions, there is a need for the exercise of discretion. There is a great deal of value in potential efficiency gains in being able to automate the easy, routine parts of the decision and free up human arbiters to deal with the discretionary aspects of a decision-making process. But it is important to recognise that the exercise of human discretion is critical in a just society, and must not be eliminated – even if this is only implemented in the appeals process.

It's also important to recognise that computers are not magic. If the policy or law is bad, all that automation will do will enable the faster application of the bad laws or bad policy, at a larger scale. A transparent, traceable decision will still deliver an unjust decision if it is the implementation of an unjust law.