Automated Decision-Making Transparency Obligation (APP 1) Issues Paper Submission

Accountability Round Table

18/06/2026

The Accountability Round Table welcomes the opportunity to make a submission to the Office of the Australian Information Commissioner's consultation on development of guidance on automated decision-making (ADM) transparency obligation.

For inquires, contact [redacted]

INTRODUCTION

About the Accountability Round Table

The Accountability Round Table (ART) is a non-partisan, civic organisation concerned with improving government accountability and democratic practices. ART is dedicated to strengthening integrity, transparency, ethical behaviour and democratic practice governments.

The need for oversight of Automated Decision-Making

Automated Decision-Making (ADM) within the Commonwealth Government is proceeding at a rapid pace, with all agencies required to have position statements about the deployment of AI by the end of June 2026. The pace is running well ahead of the development of appropriate oversight and accountability arrangements.

AI implementation processes are proceeding as if the lessons from Robodebt have all been learned and recommended remedial actions to systems of oversight and effective forward-looking arrangements to keep pace with technological developments are now in place.

Unfortunately, to the great impairment of government accountability, the Robodebt Royal Commission recommendation for the establishment of a stand-alone AI Oversight Agency to monitor, audit, and evaluate automated decision-making, has been replaced with a much-less effective initiative to expand the role of the AAT to watch out for evidence of systemic problems.

Applied to Robodebt, such limited oversight of its own relatively basic non-AI use of digitised decision making would not have diverted the responsible department from its savagely damaging administrative decisions until some years into the tragedy. We have

concerns that the deployment of AI in an NDIS context, where the use of AI for both routine and discretionary decision making is being contemplated, will replicate this tragedy.

Absence of effective oversight powers for existing oversight bodies, failing to properly fund them, and failing to establish them in the first place, seems to have become not a bug, but a feature of the increasingly degraded management of oversight processes by governments in Australia of all stripes.

The substantial frameworks the Commonwealth government has established for the deployment of AI in government services, is not matched by an equal effort to provide safeguards against the many hazards to human rights, and adequate accountability, transparency and integrity, that work in the favour of the proper execution of the government obligation to exercise a fiduciary duty in its duty to act as a guardian of the public trust (or interest).

Absent is a citizen-centric approach that is mandated to focus on assisting the humans managing the system to prioritise a 'first do no harm' approach to the human outcomes and experience of those whose lives are substantially impacted by the new AI-modified systems.

CONSULTATION QUESTIONS

The threshold question which must be addressed at the outset of this consultation since it influences all subsequent answers to this survey, is the legal standing of an AI decision. Legal authority defines whether or not an AI decision can stand on its own or whether it requires human oversight.

Legally, it is at least partially true that a "decision", cannot be legally delegated to a computer program or AI agents, as they are not legal persons nor legal entities. This principle has been established by the High Court for copyright purposes. [1] However it is not universal.

The Governance Institute of Australia also points to legal identity when setting out the legal risks and liabilities of using AI agents.

"*Unlike employees or agents in the legal sense, AI agents are* ***not distinct legal entities from their users****. Australian law is likely to treat an organisation's AI agents as part of its IT systems. This means organisations may struggle to distance themselves from the actions of their agents in the same way as they might from a rogue employee acting outside the scope of their authority.*" [2]

However no mention is made of AI legal entity status in the National AI Plan which nevertheless commits the government to upholding "*Australia's robust existing legal and*

[1] Dunlop, A., Fixler, D., & Grant Fisher. (2022, November 17). High Court denies special leave in AI inventorship case. Corrs Chambers Westgarth. Corrs Chambers Westgarth. https://www.corrs.com.au/insights/high-court-denies-special-leave-in-ai-inventorship-case

[2] Governance Institute of Australia. (2026, May). AI agents: What boards need to know about the legal risks. Governance Institute of Australia. https://www.governanceinstitute.com.au/news_media/what-boards-need-to-know-about-ai-agents/

regulatory frameworks, ensuring that established laws remain the foundation for addressing and mitigating AI-related risks". [3]

The 2024 consultation paper produced by the Attorney General's department on 'Use of automated decision-making by government' indicates that " The use of ADM to make decisions is generally limited to situations where the power or authority to take that action is contained in legislation." [4]

That paper also notes. " *The piecemeal nature of this approach was strongly criticised by the Robodebt Royal Commission. A desktop review conducted by the Attorney-General's Department has identified 46 instances of primary legislation containing provisions which authorise the use of ADM*." [5]

This Act by Act approach to regulating automated decision-making looks set to continue, although some lobbyists are pressing for blanket legality for AI decisions arguing for removal of the distinction between human and AI entity status.

Errol Price of Symmetra Pty Ltd, notes that "*Currently, Australian law and law in comparable jurisdictions, treats AI systems as tools or legal objects rather than legal subjects capable of bearing rights and obligations*." [6]

But he then goes on to argue;

"*As AI systems infiltrate more and more aspects of our lives the challenge which will have to be faced is whether it makes any legal sense to treat AI purely as a legal object. The fault lines and stresses are already apparent at a stage when AI is really only in its infancy. If AI is simply a tool or legal object then, by definition it is subject to human volition and direction. However, we know that this is not the case- even now*."

Arguing that because a practice has got out of hand, it should be allowed to continue to happen, seems to us an argument simply for poor or no regulation.

However his distinction is accurate. What is most germane to this consultation is whether an AI agent is a tool or a substitute for a human being.

It is our contention that to vest AI agents with the legal authority of equivalence to human decision making is risky in the extreme and ignores the substantial differences between the two types of decision making capacities.

---

[3] Department of Industry Science and Resources. (2026, May 7). National AI Plan. [Strategy or plan]. Department of Industry Science and Resources. Https://Www.Industry.Gov.Au/Node/95505. https://www.industry.gov.au/publications/national-ai-plan

[4] Attorney General's Department, Australian Government. (2024, November). Use of automated decision-making by government: Consultation paper. Commonwealth of Australia. p 8 https://consultations.ag.gov.au/integrity/adm/user_uploads/consultation-paper-use-of-automated-decision-making-by-government.pdf

[5] Ibid

[6] Price, E. (2026, January 19). Should AI be Recognised as a Legal Person? Legalwise; Sydney , Australia https://legalwiseseminars.com.au/insights/should-ai-be-recognised-as-a-legal-person

The questions this survey asks, rest on the assumption that ADM and AI in general will be used as a tool in the service of human decision-making, with human oversight being the ultimate arbiter.

However, the role of AI as simply a 'tool' is rapidly becoming overtaken by the capacities of agentic AI to autonomously perform entire tasks and projects. An AI agent can reason, on its own, about how to execute a human prompt for a task or project. It can also autonomously learn from its own 'experience' of doing tasks about alternative ways to carry them out.

There are several examples of these learnings going autonomously wrong. A rogue agentic AI tool "that no one had asked to have anything to do with digital money" was found to have mined cryptocurrency. [7]

A Meta AI security researcher "*told her OpenClaw AI agent to check her overstuffed email inbox and suggest what to delete or archive. The agent proceeded to run amok. It started deleting all her email in a "speed run" while ignoring her commands from her phone telling it to stop*." [8]

Furthermore, AI models have been shown to be autonomously capable of gaming their own guardrails in ways that hack social rules as well as AI's own guardrails.

A recent paper on 'reward hacking' shows how reinforcement learning in AI can allow it to make decisions on its own to find loopholes in its reinforcement learning reward functions rules. The authors show this talent can extend to what they call "societal hacking; discovering loopholes in the rules society runs on". [9]

The idea of an AI agent making an autonomous decision challenges the accountabilities of governments and government agencies in ways that using AI to automate general administrative functions does not.

The Governance Institute of Australia points out some new risks to boards and company directors which may equally apply to government agencies.

"*AI agents raise several legal risks beyond those of standard generative AI. For example, their additional autonomy increases the risk of entering into unintended transactions or contractual commitments. While agents have the potential to automate business workflows, they may not always be aware of the legal and regulatory requirements applying to those workflows. Agents are also often given access to proprietary data and internal IT systems,*

[7] Bonyhady, N. (2026, March 9). The curious case of the AI bot that went rogue and started mining crypto. The Age. https://www.theage.com.au/technology/the-curious-case-of-the-ai-bot-that-went-rogue-and-started-mining-crypto-20260308-p5o8gp.html

[8] Bort, J. (2026, February 24). A Meta AI security researcher said an OpenClaw agent ran amok on her inbox. TechCrunch. https://techcrunch.com/2026/02/23/a-meta-ai-security-researcher-said-an-openclaw-agent-ran-amok-on-her-inbox/

[9] Liu, W., Mou, X., Yan, H., Wei, Z., & He, Y. (2026). Large Language Models Hack Rewards, and Society (arXiv:2606.04075). arXiv. https://doi.org/10.48550/arXiv.2606.04075

increasing the risk of data breaches that could have significant consequences under privacy laws, cybersecurity laws and confidentiality obligations." [10]

The Governance Institute also points out that the company liabilities for agentic AI are not the same as they might be for a rogue employee.

"*Unlike employees or agents in the legal sense, AI agents are not distinct legal entities from their users. Australian law is likely to treat an organisation's AI agents as part of its IT systems. This means organisations may struggle to distance themselves from the actions of their agents in the same way as they might from a rogue employee acting outside the scope of their authority. Courts and regulators are likely to be unsympathetic towards organisations trying to avoid liability for acts or omissions of their AI agents*." [11]

These and other arguments below threaded through our responses to the OAIC questions from the "Issues Paper Submission" template will raise systemic issues which are probably beyond the scope of this OAIC consultation.

Nevertheless unless all parts of government oversight bodies and systems are aware of the Pandora's box created by human /AI equivalence, it won't just be our privacy which is in danger.

ANSWERS TO CONSULTATION QUESTIONS

Question 1

Substantially and directly related to making a decision

In your view, what are the relevant factors to enable an entity to assess whether a computer program substantially facilitates and is directly connected to a human decision-maker's decision? For example:

- *(a) Degree of reliance on the ADM system output*
- *(b) Ability and likelihood of human override over an ADM decision*
- *(c) Nature of the output (advisory vs determinative)*
- *(d)Transparency and explainability of outputs*
- *(e)Integration of ADM into decision-making workflow*
- *(f) Other*

Answers

Sub-question (a) Degree of reliance on the ADM system output.

Answer. Likert rating 5. Strongly agree that this matters. However there are some caveats

This is difficult to evaluate meaningfully. It is entirely dependent on the needs and capabilities of the individual end-user and the extent to which they rely on the output.

[10] Governance Institute of Australia. (2026, May). AI agents: What boards need to know about the legal risks. Governance Institute of Australia. https://www.governanceinstitute.com.au/news_media/what-boards-need-to-know-about-ai-agents/

[11] ibid.

Further, such guidance could lead to classification of certain ADM as substantially and directly related to making a decision even if that decision is extremely basic and non-consequential with almost zero margin for error and zero likely negative effects.

On the other hand if an inconsequential decision is material to a more consequential decision, it might be necessary to treat it as consequential as well.

For example, an output from an excel spreadsheet that calculates age from the entry of a date of birth, where the age is an essential factor in another decision, might mean that this ADM should be considered substantially and directly related to making a decision even though the original output is unlikely to have any directly adverse effects on any individuals.

Sub-question (b) Ability and likelihood of human override over an ADM decision.

Answer. Likert rating 5. Strongly agree that this, or something like it should be taken into account.

However, as with the question above, the lack of clarity of the descriptors "ability" and "likelihood" make it difficult to determine which aspects of these would be used to guide OAIC policy. "Ability" seems to refer to the extent to which the person doing the overriding has the agency to do so. That agency could be physically or socially determined. The person may have one kind of "ability" to override an ADM decision, like access to a computer, but they may simultaneously lack another kind of "ability", such as the level of workplace authority to do so.

It is impossible therefore to make a broad generalised answer to this question.

Sub-question (c) Nature of the output (advisory vs determinative).

Answer. A Likert rating is very difficult to make due to the difficulty of delimiting an advisory decision from a determinative decision. That depends on the use to which the decision is put, not what kind of an 'output' it is.

Note that this question does not define the distinction between an advisory versus a determinative output. An advisory output could become an determinative output depending on the circumstances of its eventual adoption.

This distinction needs to be explicit in the final OAIC advice to government entities since it could introduce a loophole for entities to say that an output was just ‘advisory’, and therefore not substantially and directly related to a decision, even if it was ultimately determinative. As discussed below in our answers to question 6 and above in our preamble to the questions, whether the output of ADM is deemed advisory or determinative has significant legal implications for oversight responsibility, and it also has accountability implications for government, in particular for ministerial responsibility.

To reiterate; a 'decision' (of either type - advisory or determinative) cannot be legally delegated to a computer program. [12] Unlike employees, AI agents are not distinct legal

[12] Bell and Braham 2026, op cit

entities from their users, and therefore their decisions and their mistakes are likely to be considered by courts as the liability of the organisation that uses them. [13]

It affects the discharge of ministerial responsibility, which as set out in the Howard government Guide on Key Elements of Ministerial Responsibility implies "that ministers accept two major responsibilities: first for the overall administration of their portfolios, both in terms of policy and management; and secondly for carriage in the Parliament of their accountability obligations to that institution." [14]

Note also this question is complicated by the fact that it is effectively a double barrelled question since it includes two different factors - 'advisory vs determinative decisions', to be somehow consolidated into one for rating purposes. This blending makes it difficult to answer using numerical rating alone. It is best answered in two parts.

Sub-question (c) Part 1. Advisory to a human decision.

Answer. Likert rating 4. (Agree).

This goes to arguments raised in relation to subquestion (a), 'Degree of reliance on an ADM output'. It suggests that the degree of reliance would be lower given advisory status and conditions of use than when used determinatively. Our Likert ranking response of 4, above, is predicated on ADM being used in an advisory capacity, assuming it is possible to give an ADM decision a permanent status of 'advisory'.

Sub-question (c) Part 2. Deterministic of a human decision.

Answer. Likert rating 5. Strongly agree.

Use for a determinative decision suggests that the ADM will be used to make the decision.

This is fraught with ethical, legal, and parliamentary accountability implications that are discussed in our response to question 6.

Sub-question (d) Transparency and explainability of outputs

Answer. This factor is almost impossible to rank.

1. It depends on the model design

An output, if it is a solely ADM output, will be inherently inexplicable without knowing more about its process.

That is impossible without knowing more about the actual AI model making it and the decision parameters it has been given, and the reliability of the model in terms of not hallucinating, of using reliable data for informing the decision, of not being sufficiently agentic to decide to discover loopholes in its own rules and numerous other considerations

[13] Governance Institute of Australia. (2026, May). AI agents: What boards need to know about the legal risks. Governance Institute of Australia. https://www.governanceinstitute.com.au/news_media/what-boards-need-to-know-about-ai-agents/

[14] John Howard. (1998). A Guide on Key Elements of Ministerial Responsibility Prime Minister Canberra December 1998. Commonwealth Government. https://www.accountabilityrt.org/wp-content/uploads/2009/10/The-original-Howard-Guide.pdf

which lead to AI models in general drawing unreliable conclusions from suspect or faulty data or decision parameters which are inscrutable.

While humans also experience the same kinds of difficulties in accounting for the processes which lead to human decisions so that they can be independently judged as reliable, logical and defensible (or not), we nevertheless, have existing systems which are designed to make this possible where decisions are high stakes. Judges, for instance, must make written reasons for their rulings which can be examined for their logic, their adherence to law and their general ethics.

While it is obviously desirable (and is arguable a pre-requisite for accountability in democratic systems) for outputs to be transparent and explainable, whether or not outputs are transparent or explainable has no direct relationship to whether the ADM is substantially and directly related to making a decision.

Both outputs that are very explainable and outputs that are not all explainable can both be substantially and directly related to a decision or not substantially and directly related to a decision. The Robodebt process and formula was relatively transparent and explainable, and it was substantially and directly related to making a decision.

2. It depends on who is using it

A highly opaque and unexplainable system, say the use of Copilot to make judgments on certain criteria by an inexperienced public servant could also be substantially and directly related to making a decision. [15]

While the use of Copilot is more concerning due to the lack of explainability, these examples illustrate that explainability has no impact on whether it was substantially and directly related to making a decision.

3.AI outputs are inherently unstable.

An LLM will give different answers to the same prompt question if it is asked more than once, sometimes even contradicting its previous answer. [16]

4. AI evaluations of qualitative inputs are incommensurate.

Illustrating the relatively well known methodological proposition that you can't turn a quality into a quantity (despite the best efforts of consultants to do so) the problem of incommensurability also affects AI ADM outputs.

It particularly affects the weighing of an pair of social values or social goods, as Alan McCay explains. When values are 'apples and oranges' as when judges must try to give an offender what they deserve, and also try to protect the community, or when armed conflict,

[15] For a comprehensive discussion of the ramifications of Copilot use in Australian government see; Lamchek, J., & Trieu, V.-H. (2026). Risk Regulation of Generative AI: A Case Study of Microsoft Copilot in the Australian Government. https://dro.deakin.edu.au/articles/journal_contribution/Risk_Regulation_of_Generative_AI_A_Case_Study_of_Microsoft_Copilot_in_the_Australian_Government/31069831/2

[16] Mee, T., Choi, J. J., & Baduashvili, A. (2026). Same Prompt, Different Answer: Why AI Disagrees with Itself. Journal of General Internal Medicine, 41(5), 1283–1285. https://doi.org/10.1007/s11606-025-10071-1

combatants engaging in military operations must consider incommensurables, ensuring that the civilian casualties connected with military action are not disproportionate to the military gain, the values are incommensurable and there is no common metric with which to evaluate them. How many units of beauty is worth a unit of robustness? [17]

The problem can be patched over for ADM purposes, by assigning a numerical weighting to a value, but this does not resolve the problem that humans making the same judgment as the ADM might be required to do, will be using moral reasoning and a significant degree of understanding of the social context and impact of the eventual decision in order to make it.

No amount of technological adjustment will compensate for the ADM AI not having this capacity. Assigning a numerical weighting just pushes the incommensurability problem to one side, pretending it doesn't exist. Thus the two solutions (or 'outputs') human and AI, are never equivalent.

Sub-question (e). Integration of ADM into decision-making workflow .

Answer. This question is ambiguous.

We assume it means the extent of integration into decision-making workflow.

This, as for the question above, can only be determined by knowing the conditions of the real degree of integration in a particular decision-making setting.

Sub-question (f). Other.

Answers

Sub-question (f) 'Other', Part 1. A 'decision' is not separate from the conditions that make it

A 'decision' is materially affected by the type of ADM model used, the type of oversight it has and the type of decision making required of it.

The Digital Transformation Agency website draws attention to the distinction between 'public' ADM tools and 'non-public or enterprise tools'. "*Public generative AI tools are different from non-public or enterprise tools that have been configured to meet agency data control and information security requirements. When you use public tools, your inputs and outputs may be shared with the tool provider.*" [18]

This distinction is important as non-public or enterprise tools that have been configured to meet agency requirements can be tailored to provide greater protection for agency clients from inappropriate decision-making.

However tailoring ADM also confers greater responsibility on agencies themselves to provide clear and open explanations of their own decision-making practices in designing

[17] McCay, A. (2025). Apples and oranges: AI's incommensurability problem. AI & SOCIETY, 40(7), 5607–5609. https://doi.org/10.1007/s00146-025-02317-7

[18] Digital Transformation Agency. (2026). Staff guidance on public generative AI: Using public generative AI tools safely and responsibly. Australian Government. Digital.Gov.Au. https://www.digital.gov.au/policy/ai/staff-guidance-public-generative-ai

such a tool such that it would be capable of surviving a rigorous independent commission of inquiry investigation.

1. Model

The more sophisticated an ADM model the more it will be able to manage things like context of decision. Users thinking of a decision being made by a model with sophisticated judgement capacities and high reliability (although this is only developing slowly) will make users more likely to trust it, regardless of how well it works in practice.

2. Type of oversight

Plenty of studies of organisational behaviour show that where people are responsible for checking a decision or a product for faults, they will do a worse job if they think there are numerous other people in the auditing chain who will double check their decision.

3. Type of decision required

If, as now seems likely for NDIS decisions, ADM is employed to make complex discretionary decisions it is more likely not to be queried. In ordinary human organisations simple decisions are made by people at the bottom end of the organisation and complex decisions are handed upward through layers of organisational managers until they reach someone who must decide them despite their complexity. Usually those managers have higher levels of organisational accountability. It is unlikely that the decision of a model sufficiently sophisticated, as to be able to handle discretionary decisions, will be easy to scrutinise for accountability, especially if it is able to learn from the results of its own reinforcement learning.

Sub-question (f) 'Other', Part 2. Is a 'decision' visible accessible and auditable? Implications for FOI and oversight bodies

In order for government to be transparent and accountable, information on decisions must be accessible. In a speech given by Anne Twomey at Melbourne University in May of this year, she focuses on inadequate documentation of uses of public money in government. Adequate management of what should be public information, she argues, needs three things; "First, that records are made. Second, that they are kept. Third, that access is granted to them, without undue limitations, at the appropriate time." [19]

In her speech, her main example of evasion of statutory documentation generation and preservation requirements was the behaviour of the Berejiklian government in NSW.

In that case it was email which was used as a means of evading documentation.

"*While the decision-making power for the Stronger Communities Grant Scheme was vested in the Minister for Local Government, all but two of the funding decisions were made by the Premier and Deputy Premier. The projects they chose and the relevant amounts were then sent by email from ministerial advisers to the Office of Local Government. That Office then*

[19] Twomey, A. (2026, May 27). 'No Paper and No Trail – Government Unaccountability in the Digital Age' Speech delivered at the 2026 Jim Carlton Integrity Lecture, Accountability Round Table, University of Melbourne at Melbourne University Law School.

paid out hundreds of millions with no formal ministerial authorisation or any documentation about the basis for the selection of most of the grants." [20]

Then the documents were destroyed. "The hard copies were physically shredded and the electronic copies were digitally deleted." [21]

Using then deleting email is a relatively primitive form of digital record evasion, and in this case it didn't work, as the inquiry looking into it issued an order for the documents in the backup electronic files.

These days there are so many platforms, so many secure communication channels and so many self erasing messaging apps that this kind of digital resurrection is nearly impossible.

Using them is not necessarily a form of deliberate evasion, it can be a matter of simply shifting private personal practice to a government work place.

The same issues apply to use of ADM. Government agency staff may now commonly have access to AI chatbots for decision-making. This will likely be considered 'not integrated' even though these may be informally used to make decisions.

How much harder, however to ensure records are made and kept and are accessible if the decision-maker is agentic AI.

The following powers of Claude Cowork from its January 2026 release, are quoted below in full. It's the easiest way to make clear the complexity and reach of agentic AI and the difficulty of record keeping.

Should AI prompts by humans be part of record keeping? Should the AI be banned from executing any kind of "hide" or "destroy" commands? How?

"*How is using Cowork different from a regular conversation? In Cowork, you give Claude access to a folder of your choosing on your computer. Claude can then read, edit, or create files in that folder. It can, for example, re-organize your downloads by sorting and renaming each file, create a new spreadsheet with a list of expenses from a pile of screenshots, or produce a first draft of a report from your scattered notes.*

In Cowork, Claude completes work like this with much more agency than you'd see in a regular conversation. Once you've set it a task, Claude will make a plan and steadily complete it, while looping you in on what it's up to. If you've used Claude Code, this will feel familiar—Cowork is built on the very same foundations. This means Cowork can take on many of the same tasks that Claude Code can handle, but in a more approachable form for non-coding tasks.

When you've mastered the basics, you can make Cowork more powerful still. Claude can use your existing connectors, which link Claude to external information, and in Cowork we've added an initial set of skills that improve Claude's ability to create documents, presentations,

[20] Ibid
[21] Ibid

and other files. If you pair Cowork with Claude in Chrome, Claude can complete tasks that require browser access, too." [22]

A related problem is the increasing ease of "vibe coding".[23] Vibe coding has a very low standard of technical expertise needed to master it. This bar will only get lower as AI coding models improve. If public servants and other government officials can code and build their own specialist AI models and use them in order to avoid scrutiny, then what hope is there for recordkeeping or oversight? This is the Berejiklian government email problem on steroids.

Question 2

Substantially and directly related to making a decision

Fictional Edge Case: Generative AI with human oversight

The National Aged-Care Insurance Agency (NACIA) makes decisions about eligibility for elderly people trying to access life-changing aged-care support. To make eligibility assessments quicker, NACIA staff use a generative AI chatbot, GPTea, to summarise candidate profiles and receive recommendations about eligibility. NACIA's human staff always have oversight and make the final decision about eligibility based on the recommendation provided by the chatbot.

(a) Would you consider NACIA's described uses of GPTea to be substantially and directly related to making decisions? (Please provide yes/no)
(b) Can you provide your reasons for your response?

Answers:

Embedded in the current approach to government decision oversight is an assumption that humans are capable of it, and that they will remain more capable than AI oversight of ADM. There is a very good chance that the pressure will be on governments very soon to replace human oversight with AI oversight for reasons of efficiency. This is despite the ethical dubiousness of machines making decisions on how humans should be governed and other machines then deciding if they are right, and the aforementioned legal constraint that neither are legal persons nor legal entities.

The efficiency problems are growing as government service management becomes broader and more complex and a greater number of third party contractors are engaged in design and delivery.

---

[22] Anthropic. (2026, January 12). Introducing Cowork. Claude. https://claude.com/blog/cowork-research-preview

[23] Yildirim, E. (2026, February 23). An AI Thought Experiment on Substack Is Sending the Stock Market Spiraling. Gizmodo. https://gizmodo.com/an-ai-thought-experiment-on-substack-is-sending-the-stock-market-spiraling-2000725601

As a present response to volume of work, Copilot is already commonly used to summarise documents in the Australian Public Service, [24] allowed by the fact that the Australian federal government's end-user guidance only prohibits the use of AI as the final decision-maker. [25]

There are (at least) two problems with humans overseeing AI decisions.

1. Humans can't reliably detect AI misjudgements and mistakes.

2. Top level AI can detect AI mistakes and human mistakes better than humans.

However society is better adapted to creating accountabilities for humans than it is for investigating AI failures and creating accountabilities for them. Even Anthropic's Mythos, a general-purpose language model which can perform all AI tasks "*but it is strikingly capable at computer security tasks*" has at the end, human oversight. [26]

Even if that model and its public use offshoot, Fable are currently disabled by a political decision of President Trump, it is only a matter of time before other more politically favoured AI companies catch up with Mythos' capabilities and they are extended to work with error management of decisions using language rather than code error detection. As an aside, it doesn't take much imagination to work out that the government definition of "decision error" in today's political climate, might have a down side.

Explanation 1. Humans can't reliably detect AI misjudgements and mistakes because these decisions are inscrutable, subject to AI hallucinations, and can be biased to "majority" profiles (and therefore miss normal diversity) depending on the frequency and type of "normal" data they have been trained on.

Those first three weaknesses are well known. However there are some lesser known weaknesses to AI especially in a sensitive social context such as disability or aged care decisions.

As an example, the NDIS Bill goes further than any other government regulation into extending the idea of automated decision-making into the making of discretionary decisions. [27]

This is highly problematic. Few AI models, except top of the range models, are capable of "context" in decision-making. The default is more narrowly defined and formulaic, like Robodebt.

Second, it is unclear what model of context an AI model would use for disability decisions if it actually was context aware. It could be a social model of disability which looked at how

[24] Lamchek, J., & Trieu, V.-H. (2026). Risk Regulation of Generative AI: A Case Study of Microsoft Copilot in the Australian Government. Op cit p 29

[25] Ibid p 24, quoting the Australian Government Transformation Agency

[26] Carlini, N., Cheng, N., Lucas, K., Moore, M., Nasr, M., Prabhushankar, V., Xiao, W., Angulu, H., Asher, E. B., Bow, J., Bradwell, K., Buchanan, B., Forsythe, D., Freeman, D., Gaynor, A., Ge, X., Graham, L., Guru, K., Lakhani, H., … Troy, K. (2026, April 7). Claude Mythos Preview. Red.Anthropic.Com. https://red.anthropic.com/2026/mythos-preview/

[27] Barlow, K. (2026, June 6). Exclusive: Unprecedented automation in NDIS decisions. The Saturday Paper, (603). https://www.thesaturdaypaper.com.au/news/health/2026/06/06/exclusive-unprecedented-automation-ndis-decisions

people understood their own level of disability, or it could be an environmental model of disability which confined itself to 'safety' considerations. [28]

Explanation 2. Top level AI can detect AI mistakes and human mistakes better than humans. Anthropic's Mythos demonstrates a first example of an AI model that outpaces human ability to detect coding vulnerabilities in software, whether written by humans or by AI agents that can code. Estimates of how much coding is now done by AI agents varies widely, but the trajectory appears to be steeply upward. Claude Mythos significantly increased the number of vulnerabilities detected in software from those discovered by prior 'red team' security assessments done by humans.[29]

It is not difficult to predict that this capacity could be extended to looking for vulnerabilities in any decision or system, human or machine.

However, as with human decision-making, this capacity to can be used for socially positive and socially negative purposes as the following quote from Anthropic's assessment of Claude Mythos capabilities attests;

"*There are only so many classes of vulnerabilities, and through a combination of intelligence, encyclopedic knowledge of prior bugs, and an ability to be far more thorough and diligent than any human can be (though they are still imperfect!), language models are now remarkably efficient vulnerability detection and exploitation machines*." [30]

The real question here is how to create a solid system of accountabilities **that does not allow social responsibility - an utterly human responsibility - to be delegated to a non human**. Without human responsibility, political responsibility becomes meaningless. [31]

Question 3

Meaning of significantly affect rights or interests

What factors do you consider increase the likelihood that a decision could affect an individual's rights or interests? For example:

- *Sensitive information*
- *Vulnerable persons*
- *Intrusive practices*
- *Financial outcomes*
- *Other*

Answers.

[28] Kim, E., Tanase, I., & Mallon, C. (2026). DisaBench: A Participatory Evaluation Framework for Disability Harms in Language Models (arXiv:2605.12702). arXiv. https://doi.org/10.48550/arXiv.2605.12702 and see also discussion on p12 of this submission of different social frames for the idea of disability.
[29] Carlini et al, op cit
[30] Ibid
[31] See answers to Question 6 below, for more on government responsibility

This risk list above is largely limited to informational risks. Deficit definitions can be framed in a multitude of ways, each one of which will generate a different risk list.

If the deficit is understood to be an environmental deficit its risks will be framed in terms of injury and physical environment design.

If the deficit is framed as a social deficit its risks will be framed in terms of social contact and capacity and opportunity for social agency.

If the deficit is framed as a power deficit, its risks will be framed as problems of discrimination, invisibility and lack of access to the social and political power necessary to change the conditions of living and acting in society.

If the deficit is framed as a "rights" deficit, its risks will be framed as those of exploitation by others of government benefits, risks of abuse by carers, absence of pathways for redress, and lack of fundamental human rights.

Every framing will generate its own list of risks. The list of risks provided for this question are fundamentally a framing in terms of misuse or abuse of client information. But introducing ADM will inevitably introduce a number of other risk frames that the ADM fails to register as key decision-making factors.

Question 4

Meaning of significantly affect rights or interests

What classes of person do you consider to be vulnerable for the purposes of considering a decision's expected effect on a person's rights or interests, and why?

Answers

Any person whose life opportunities can be defined in terms of a deficit model that explains underperformance in terms of a specific shortcoming.

Deficit models are robustly critiqued by social scientists as an inadequate descriptor or diagnostic tool, but are still routinely used to provide protection, services or benefits to those people who are deemed to have them. The result is usually an inadequate package which reflects the narrow definition of their "deficit".

Question 5

Meaning of significantly affect rights or interests

What do you consider to be a significant service or support, for the purposes of considering whether a decision affects access to a significant service or support? For example:

- *physical assistance/support*
- *financial assistance*
- *access to educational services*

- *'in home services'*
- *access to essential banking and credit services*
- *access to telecommunications*
- *access to essential utilities*
- *other*

Answers

As discussed above, the framing of the deficit sets up the framing of the risk. The framing of the risk will therefore set up the expected remedy to the problem as it does above.

This question frames the **rights or interests to** 'service or support' atomistically.

If the question was considered systemically, "*for the purposes of considering whether a decision affects access to a significant service or support*", a possible affected support might be a broad societal one of providing for the human rights pertaining to any 'deficit' whether determined by age, ability, race, power financial capacity, environment, geography or any other similar inequity, and the capacity to pursue redress for equity shortcomings.

A possible interpretation of such a right might be the right to be genuinely, (not formulaically) individually consulted about the supports needed by each individual to make their lives workable.

AI could be a significant factor in **providing** assistive services directly to the user. However, as this article argues, it would best assist as a service, if the design process incorporated a guiding framework to center lived disability experiences from the perspective of the user. [32]

Question 6

Meaning of significantly affect rights or interests

Are there any other legal frameworks or policies that the OAIC should have regard to in considering the meaning of 'rights' and 'interest'?

Answers

The 2024 consultation on Automated Decision Making by the Attorney General's Department invoked some salient principles of matters that might affect rights or interests by use of ADM.

• ethical considerations and principles of administrative law must be taken into account in the design of automated decision-making processes, and there should be proper oversight of these arrangements

• legislative reforms should ensure clear review pathways for those affected by decisions, and

[32] Tang, X., Lin, T., Li, J., & Wu, S. (2026). Cripping AI: Reimagining AI Through Lived Disability Experiences. arXiv. https://doi.org/10.1145/3805689.3806744

• transparency should extend to both the use of ADM, and how decision-making processes operate. [33]

Attorney General's Department, Australian Government. (2024, November). Use of automated decision-making by government: Consultation paper. Commonwealth of Australia. https://consultations.ag.gov.au/integrity/adm/user_uploads/consultation-paper-use-of-automated-decision-making-by-government.pdf

There are some further issues of how the application of ADM may have some future effects on social and cultural understanding of both 'disability" and "decision-making" to the detriment of service users and to the idea of government responsibility.

<u>The introduction of Algorithmic authority.</u>

The effects of normative changes - custom and practice issues - can have as much impact on social constraints and oversight as legal and ethical principles or regulatory rule changes. The effect is recursive. Principle and rules can affect normative behaviour and normative behaviour can affect principles and rules.

So for instance Andy Phippen argues that the use of AI tools "where automated systems materially structure the reasoning underpinning findings of incapacity, and human oversight becomes merely formal" the risk increases that rights and rules may become eroded. "Drawing on the concept of systemic drift, (he) contends that the principle danger lies not in machine autonomy, but in the gradual normalisation of algorithmic authority within legally consequential decision-making processes." [34]

The slow locking in of the understanding that machines can make decisions for humans is also a case in point.

Unthinkable once, the introduction to the Attorney General's 'Use of automated decision-making by government: Consultation paper' shows how incremental legal change creates a changed normative environment for ADM.

"*The use of automated decision-making (ADM) by government is long standing practice. For example, section 6A of the Social Security (Administration) Act 1999, which provides that the 'Secretary may arrange for use of computer programs to make decisions', was inserted into that Act nearly 25 years ago in 2001. ADM provisions have since been adopted and currently operate across major Commonwealth service delivery contexts, including social services, migration, biosecurity, aged care and veterans' entitlements*." [35]

The normalising process for automated decision-making has been much longer than most people imagine. This long embedding of it as an accepted social practice will make any significant reforms or controls very difficult.

---

[33] Attorney General's Department, Australian Government. (2024, November). Use of automated decision-making by government: Consultation paper. Commonwealth of Australia. https://consultations.ag.gov.au/integrity/adm/user_uploads/consultation-paper-use-of-automated-decision-making-by-government.pdf

[34] Andy Phippen. (2026). The Banality of Automation: How ordinary Administrative AI Potentially Threatens Extraordinary Rights. Computer and Telecommunications Law Review.

[35] Attorney General's Department, Australian Government. (2024, November). op cit

Democratic erosion and ministerial responsibility

A norm that has been steadily eroded over the last decade or more by general refusal to acknowledge it exists, rather than by ADM, is the idea of ministerial responsibility. The principle is that the Minister stands as the ultimate source of accountability for the actions of his or her departments. As this idea has slipped away, so has systematic oversight. The way has now been normalised for Andy Phippen's "algorithmic authority" to take over, and fill the gap.

That this is questionable for the proper operation of government or indeed, democracy is clear in the arguments Justice Bell uses to defend the courts from similar erosion, noted in the preamble to this document to our remarks.

In his speech, Bell defends judicial independence as a way to separate judicial decisions from influence by the executive. In this respect it is possible to see an analogy with ministerial responsibility. It too is a way to uphold the accountability of the minister's own portfolio of agencies by separating the minister from the internal politics and self preservation of the public service, by insisting he or she account for its mistakes. Parliament is ultimately the primary source of accountability for government decision-making and it should be its job to hold ministers responsible for the decisions of their departments. [36]

The proper operating of government is also a factor in the extent to which use of AI and especially AI agents, **significantly affect rights or interests,** particularly if those agents are making sophisticated discretionary decisions if such use interferes with ministerial responsibility.

Ministers have a duty to seek out and be informed of the decisions that are made within their ministerial portfolios which have an impact on their management of public money, their obligation to be able to monitor and explain the carrying out of government work by their departments, and to weed out corruption or wrongful behaviour in all matters for which they have responsibility. For ministers not to be fully apprised of departmental decisions puts them into a position where they may find themselves misleading parliament, or being embarrassed by journalists when they don't know what their department is doing, or how much it is spending and on what, and whether or not that is a matter of corruption. Worse, the misuse of public funds can give rise to the criminal offence of misconduct in public office.

These informational transparency and accuracy obligations may well be significantly affected by the introduction to government services of automated decision-making of any kind.

State regulatory capture

There is also a much larger problem created by handing over more and more government and public service tasks to AI models and entrenching ADM as a critical administrative service. It stands a good chance of creating dependence of Australia government for its internal administrative and institutional infrastructure, on a small number of very rich and powerful tech company owners.

[36] Bell and Braham 2026, op cit

The resistance of these owners to any forms of regulation has been amply demonstrated in Australia via the problems of the AI Safety Act which has been met with fierce resistance and much lobbying by social media company owners. The reaction of Facebook in particular to being asked to pay for the news items that kept users on its platform is another instance of regulatory resistance, as is the swift about face on proper regulation of AI through a purpose build Act in Australia in mid 2025, after Trump was duchessed by these same company owners and their catch cry that "regulation stifles innovation" spread to the Australian government via the report of the Productivity Commission inquiry of December 2025. [37]

Question 7

Meaning of significantly affect rights or interests

Fictional Edge Case: Differential pricing

E-commerce company, Daintree, allows subscribers to order books to their home. To make an account, Marco was required to disclose his address, being the luxurious Capital Crescent in Sydney CBD. Marco is surprised that his desired book, The Art of Inheritance, was priced at $100 when his book-club buddy, Charles, from a country town, paid only $35.

Marco is shocked to read an article reporting that Daintree uses differential pricing to inflate prices for individuals from wealthy postcodes. Marco is a freelance artist, so he does not have enough disposable income to purchase the book.

OAIC remarks

(a) Restricting access to financial products or healthcare would clearly be a decision which significantly affects the interests of an individual.

In the above Edge Case, would you consider the inflation of prices for a consumer product to have significantly affected the interests of Marco?

If yes, at what amount would differential pricing be considered to have significantly affected the interests of an individual?

5% price difference, 20% price difference, or other (please specify amount)

(b) In judging whether differential pricing significantly affects rights or interests, how important would you consider the following factors on a scale from 1 to 5 (with 1 being not important and 5 being very important):

- *The dollar amount of differential pricing*
- *The significance of the product or service being purchased*

Answers

[37] Australian Government Productivity Commission. (2025, December). Harnessing data and digital technology—Public inquiry. Commonwealth Parliament Australia. https://www.pc.gov.au/inquiries-and-research/data-digital/

The most significant effect of differential pricing is not the differential cost, although that is clearly unfair.

It is the effect of differential pricing on the manipulation of consumer behaviour to the benefit of the seller. It is done by algorithms **learning the decision trigger points of the individual target** in order to optimise their response in favour of the seller, or service contractor, in the case of use of ADM for disability resource distribution.

If a seller of computers, for instance, knew via its customer profile database, that a purchaser had a disability, and that the computer might be subsidised because it fits the definition of a medical device under section 41BD of the Therapeutic Goods Act 1989 [38] they may decide to increase the price of the computer to that customer to absorb some or all of the subsidy.

This could materially work to advantage a third party contractor to perpetrate an AI driven form of rort on clients and on state subsidies.

To be clear, this potential problem is not caused by ADM being used by a government service to make decisions about client eligibility for services, working against the interests of the client. In this case, it is most likely to be a private supplier of medical equipment using a sophisticated customer management database to identify vulnerable customers.

It is in part a problem because the Therapeutic Goods Administration defines a medical device by its purpose.

"*Australia's medical device regulatory framework is technology agnostic which means we regulate products based on their intended purpose, not the technology they use. The intended purpose – what the product is used for – is defined by the manufacturer and will determine whether the product meets the definition of a medical device under the Act*". [39]

Relying on a definition of 'intended purpose' by a manufacturer in whose interests it is that sellers sell more of their product seems like a real legislative weakness that can be exploited by AI.

The inflation of prices for a consumer product by seller use of AI would have significantly affected the interests of Marco, and the interests of the government providing the subsidy.

Question 8

Meaning of making a decision

Fictional Edge Case: Discriminatory targeted job ad

[38] Therapeutic Goods Administration (TGA). (2025, May 6). Therapeutic Goods Act 1989 | Therapeutic Goods Administration (TGA) [Text]. Therapeutic Goods Administration (TGA). https://www.tga.gov.au/resources/legislation/therapeutic-goods-act-1989

[39] Therapeutic Goods Administration (TGA). (2024, May 7). Artificial intelligence (AI) and medical device software regulation | Therapeutic Goods Administration (TGA) [Text]. Therapeutic Goods Administration (TGA). https://www.tga.gov.au/products/medical-devices/software-and-artificial-intelligence-ai/manufacturing/artificial-intelligence-ai-and-medical-device-software-regulation

Engineering firm, Manscon, is hiring for graduate engineers. To try to target the best candidates, Manscon advertises open roles on job site, Opaque Window. Opaque Window algorithmically promotes this job posting to candidates it thinks are best suited. One metric it uses is gender of account holders. Since most engineers at Manscon are male, Opaque Window prioritises promoting the job to male engineering graduates.

Rafqa is a female engineering graduate searching for a job on Opaque Window. She does not receive the Manscon job ad, which limits her employment options.

OAIC remarks

Should it be considered a 'decision' that Rafqa did not receive the Manscon job ad?

Please provide a Yes/No response

Can you provide your reasons for your response?

Answers

Yes, it's a decision under the present ad hoc governance systems in which AI and ADM decisions are executed and where no distinction is made between a human decision and a machine decision. It's a decision of omission, not commission, but is still a decision, and not a random act of nature, since it has 'intention' to make a candidate selection. But it's a biased decision.

The question goes to the very well known propensity of algorithms and AI to be biased. This machine bias can be unintentional because of the limited nature of the data sets it uses, or ill considered instructions given to the AI model.

Or it can be entirely intentional, as AI models like Elon Musk's Grok have shown. Musk regularly and publically alters the Grok algorithm to promote his political views through the answers given by Grok, and the targetting of people susceptible to them.

What guardrails do we have in place for preventing some government of the future from deciding to eliminate all "woke" use of ADM in provision of government services?

Question 9

Meaning of 'arranged for'

Are there any scenarios relevant to the meaning of 'arranged for' that you consider require guidance? If so, why?

Can you provide your reasons for your response?

Answers

This is a very difficult question to answer since the term "arranged for" here, is presented context free. Context changes meaning.

The best attempt we can make is to give the phrase the context of an AI agent "arranging for" something to happen.

Does this have equivalence with "making a decision"? With respect to appropriate transparency and accountability, all the caveats which apply to "a decision" would appear to also apply to an action by agentic AI.

Question 10

Extent of disclosure

When making disclosures about ADM, the OAIC's view is that entities must ensure that meaningful and accessible information is provided to consumers. In doing so, the OAIC considers that entities should strike an appropriate balance to ensure that disclosures are:

- *clearly articulated in plain language and easy to understand;*
- *structured to enable consumers to request further information, where required;*
- *appropriately tailored – sufficiently specific to be meaningful, while avoiding overwhelming levels of detail;*
- *organised so that similar information is grouped in a logical manner; and*
- *framed in a way that allows the information and the decision to be challenged or contested.*

Having regard to the above considerations, is there anything additional that entities should disclose to provide effective and meaningful disclosures about ADM?

Answers

We strongly recommend that entities should disclose information about **possible recourse** where they believe there has been a breach of privacy or a discriminatory decision made.

The OAIC Issues Paper states [40] that the obligation will allow people to take further action if there has been a breach of personal privacy, but so far we have seen no evidence as to how people will actually be able to take action, particularly those that have limited levels of literacy about privacy and ADM which would arguably be a majority of Australians.

Other Matters

Meaning of computer program

We welcome the guidance outlined by the OAIC for the meaning of computer program to be interpreted broadly.

We would like to emphasise that non-deterministic, machine learning ADM systems are a whole other order of magnitude compared to ordinary algorithms, and should arguably be considered as a separate class of computer system.

[40] OAIC Issues Paper p 3

The Senate Select Committee on Adopting Artificial Intelligence commented, *"…given the key risks of AI, the application of AI to ADM systems raises significant issues"* [41] which indicates the step-change in consequences of more 'intelligent' ADM:

Guidance should provide that ADM that incorporates any form of non-deterministic or discretionary decision-making should be distinguished, and special consideration must be given in assessing and disclosing the use of such ADM. This consideration should include acknowledgment and consideration of the above factors in a privacy policy, and other factors outlined later in this submission related to non-deterministic ADM, including but not limited to the inability to know the extent of influence such ADM will have on any particular end-user within an organisation.

[41] Select Committee on Adopting Artificial Intelligence (AI). (2024). Select Committee on Adopting Artificial Intelligence (AI) Final Report [Text]. Commonwealth Parliament. (Australia). https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Adopting_Artificial_Intelligence_AI/AdoptingAI/Report page 137.