[redacted] | Governance and risk adviser | [redacted]

Submission to the Office of the Australian Information Commissioner

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

11 June 2026

Submitted via: privacyreformtranche1@oaic.gov.au

About this submission

[redacted].

The framework is a governance accountability standard that defines the conditions under which reliance on consequential AI-enabled decisions and actions can be justified, defended and stood behind - not as proof of legality, but as the evidentiary condition for defensible organisational reliance.

The examples and implementation observations in this submission draw on direct experience working with APRA-regulated entities navigating governance challenges in operational settings.

This submission addresses Questions 1, 2, 9 and 10. Questions 3 to 8 are not addressed.

The submission draws on the Warranted AI framework to offer a perspective that is intended to complement, rather than duplicate, the technical and legal analysis that other submissions are likely to provide.

The distinctive contribution is that transparency is a necessary condition for accountability, but not a sufficient one. The deeper test is whether governance actually operated when an AI system shaped a decision - not whether governance was documented, intended, approved or described. **Governance intent** describes what the organisation meant to apply. **Governance in Effect** means governance shown to have operated in practice - as decisions were generated, actions were taken, consequences emerged and risk profile changed. The gap between those two concepts is where this submission sits.

Question 1 - Substantially and directly related to making a decision

The model (computer program) is not the mechanism

The factors listed in the Issues Paper - degree of reliance, likelihood of human override, nature of output, transparency and explainability, integration into decision-making workflow - are individually useful. But they share a structural limitation: each treats the computer program as if it were the mechanism under assessment. The Warranted AI framework's first condition, Mechanism Transparency, exposes why that framing is insufficient.

Mechanism Transparency asks whether the organisation knows the decision-and-action mechanism on which it relied. The condition is not satisfied by merely knowing that an AI model was used, nor by having a high-level process map or a vendor description. The relevant mechanism includes the broader arrangement that produced or materially shaped the decision or action. This extends beyond the model itself to encompass the data the model used, the instructions and context it received, the workflow in which it operated, the human steps and escalation paths involved, and the downstream systems that acted on its output.

If the 'substantially and directly related' test is applied only to the computer program in isolation, entities may satisfy the test on paper while the actual mechanism of reliance remains opaque. The question the Guidance should direct entities to ask is not only: does this computer program substantially and directly relate to a decision? It should also be: does our assessment of that computer program account for the full mechanism through which the decision is produced?

What this looks like in a regulated entity's architecture

In a credit provider, the mechanism is not the credit scoring model alone. It is:

- the data pipeline feeding the model,
- the threshold rules that convert its output into an approval, a decline, or a referral for manual review,
- the case management workflow that determines the conditions under which a human reviewer can exercise override, and
- the downstream origination system that executes the credit action.

Each component is distinct and consequential. Assessing substantiality at the model level - while the consequential structural work happens across the surrounding architecture - locates the transparency obligation in the wrong place.

The same pattern appears in claims assessment, Anti-Money Laundering alert triage, hardship routing and complaints prioritisation. In each case, an assessment that focuses on the scoring model or recommendation engine, without accounting for the threshold rules, workflow constraints, override conditions and downstream execution, understates both the substantiality of the arrangement and the scope of what must be disclosed.

This is not a failure of governance intent. Regulated entities typically have model risk management policies that document the scoring model. The problem is that the governance artefact - the model validation report, the use-case approval - sits at a different level of abstraction than the mechanism that actually produced the decision. The Guidance should encourage entities to test whether their transparency assessments reach the full mechanism, not only the model within it.

Ranking of listed factors

Integration of ADM into decision-making workflow - 5 (Strongly agree)

The most important factor because it directs attention to the mechanism as a whole. An entity assessing integration is forced to look beyond the model to the data pipeline, the threshold rules, the override conditions and the downstream execution path. It is the factor most likely to capture genuine substantiality of reliance.

Degree of reliance on ADM system output - 5 (Strongly agree)

The primary test of substantiality. An output that decision-makers routinely accept without independent verification is substantively different from one that is cross-checked against other evidence. Reliance in practice, not nominal label, is the relevant test.

Nature of the output (advisory vs determinative) - 4 (Agree)

Relevant but not determinative. An output labelled 'advisory' may in practice be treated as determinative where reviewers lack the time, information or authority to exercise independent judgment. The label should not substitute for an assessment of actual practice.

Transparency and explainability of outputs - 4 (Agree)

Relevant to whether human oversight is capable of being genuine rather than nominal. If an output cannot be understood or interrogated by the human reviewer, its advisory label is unlikely to reflect actual practice.

Ability and likelihood of human override - 3 (Neutral)

Capability to override is not the same as the exercise of override. An entity may technically permit override while structuring the workflow, time pressure and performance incentives in a way that makes override unlikely. See also the submission's response to Question 2.

Additional factor: mechanism scope

The Guidance should explicitly direct entities to consider whether the system under review is the whole mechanism or one component within a broader mechanism. An entity that has arranged for multiple components - a scoring model, a threshold rules engine and a case management system - each of which individually may appear marginal, may together constitute an arrangement that substantially and directly shapes the decision. Assessment at the component level, rather than the mechanism level, allows structural disaggregation of what is, in practice, a compound reliance chain.

Question 2 - Fictional Edge Case: Generative AI with human oversight (NACIA / GPTea)

Response: Yes

GPTea's described use should be considered substantially and directly related to making decisions. The Explanatory Memorandum states that a computer program is substantially related to a decision where it is a key factor in facilitating the human's decision-making. A generative AI chatbot that summarises candidate profiles and provides eligibility recommendations plainly meets that description. The recommendation is what the human decision-maker acts upon. NACIA staff always retaining final decision authority does not change the character of the relationship between the program and the decision.

Human oversight is not itself a warrant

The more important point for Guidance purposes is what the NACIA scenario reveals about oversight. The Warranted AI framework's sixth condition, Supervisory Independence, addresses this directly. Supervisory Independence asks whether the system could be observed, challenged, interrupted or overridden through a capability not dependent on the system's own self-assessment.

This condition matters because a confidence score, model-generated summary, internal alert or system log may assist supervision. But those artefacts do not establish independent oversight if the organisation has no separate capacity to challenge, interrupt or override the system when reliance becomes unsafe. Recent research in frontier AI systems reinforces the same distinction: detection and escalation capabilities are improving, but interruption and independent override remain materially weaker - a gap that nominal human-in-the-loop arrangements do not close.

The condition fails where a human reviewer lacks the authority, information, time or practical ability to challenge the recommendation - even if that reviewer nominally makes the final call. In the NACIA scenario, the critical governance question is not whether a human makes the final decision. It is whether NACIA staff have an independent basis on which to challenge GPTea's recommendation - or whether, in practice, they defer to it because they cannot evaluate or contest it effectively.

The oversight failure is structurally worse in commercial regulated entities

NACIA, as a government agency, operates within a public law accountability framework. Its decisions can be challenged through administrative review, merits review and ombudsman mechanisms. Those accountability structures create external pressure that at least partially compensates for weak internal oversight. An APRA-regulated financial entity processing thousands of credit, claims or membership decisions daily does not have the same compensating structure - and faces governance pressures that systematically work against substantive oversight.

Volume pressure compresses individual review time. Performance incentives that reward throughput, not exception handling, reduce the practical likelihood of override. The information asymmetry between a reviewer and a model that processes hundreds of variables is typically far greater in a financial services environment than in a government agency context. Staff may lack access to the underlying model logic entirely, seeing only a recommendation and a confidence level. In that environment, human oversight is structurally

likely to be nominal regardless of what the governance policy says - not because of any individual failure, but because the conditions for substantive review have not been created.

The consequence pattern for decisions of this kind is identifiable: individuals who should have been approved are declined; cases that warrant human review are not escalated; exceptions that require judgment are absorbed into the automated recommendation without scrutiny. The governance exposure is not the individual decision that is later challenged. It is the entire class of decisions in which the AI recommendation shaped the outcome without the quality of oversight the entity's governance documentation implies.

The Guidance should not allow entities to satisfy the ADM obligation merely by inserting a human decision-maker at the end of an automated recommendation process. Where human oversight is part of an entity's ADM governance claim, the transparency obligation should extend to describing the quality of that oversight - whether it is positioned to affect reliance, not merely to record it.

Question 9 - Meaning of 'arranged for'

'Arranged for' is an accountability preservation question

'Arranged for' should be understood as tracing accountability to the right place. It is not a question about technical operation. It is a question about who must be able to stand behind the reliance chain when things go wrong, when regulators ask or when an affected individual seeks to understand what happened.

Warranted AI's third condition, Accountability Through Delegation, holds that accountability for AI-enabled decisions and actions must survive delegation. An organisation may distribute technical operation, monitoring, execution and support to a vendor, a managed service provider, a cloud provider, a model supplier or a fourth-party dependency. That delegation is legitimate and often necessary. But it must not dissolve accountability.

An entity that has arranged for a third-party AI program to screen loan applications, assess insurance claims or triage service requests has not merely enabled another entity to operate a system. It has arranged for that program to make or materially shape decisions about its own customers, using personal information it holds and is responsible for. The accountability for the decision pathway remains with the entity that arranged it - not only with the entity that operates the program.

The failure pattern is familiar: the business function says the model was operated by technology; technology says the use case was owned by the business; the vendor says the system operated within specification; the reviewer says they relied on the output; the downstream team says it merely executed the workflow. In that structure, many parties carry responsibility for a part of the chain, but no one is positioned to stand behind reliance on the decision or action as a whole. That is the accountability gap that 'arranged for' must be interpreted to prevent.

Three scenarios that require particular guidance

1. Layered arrangement: Many entities arrange for AI-enabled decisions through multiple intermediaries - a platform vendor whose product incorporates a model supplied by a third party, which draws on data processed by another provider. The question is not only whether the entity arranged for a computer program. It is whether the entity has preserved accountability across the full chain of delegation. Where the chain is long, an entity cannot satisfy an accountability-preservation standard by pointing only to its immediate vendor relationship. The Guidance should make clear that the obligation follows the arrangement through its dependencies.

2. Platform-mediated arrangement: An entity may use a general-purpose software product - a CRM, an origination system, a member engagement platform - that includes AI features where that functionality was not the primary reason for procurement and may not have been considered at the time of acquisition. The entity has, nonetheless, arranged for the program to make or materially shape decisions. The Guidance should make clear that the obligation follows the arrangement, not the intent at the time of procurement. Entities that do not actively audit the AI features embedded in their existing platform environment risk non-compliance by omission.

3. CPS 230 material service provider arrangements: Many APRA-regulated entities now operate under the material service provider framework introduced by CPS 230. They

have procured platforms - core banking systems, origination tools, claims management systems, member engagement platforms - that may have incorporated AI features since the time of procurement, often without deliberate governance decisions about those features at the point of acquisition. In many cases, the entity has arranged for ADM within the meaning of APP 1.7. It has not recognised that it has done so.

There is a practical coherence opportunity here that the Guidance should address. An APRA-regulated entity's material service provider register under CPS 230 may already document some of the same third-party arrangements that constitute 'arranged for' under APP 1. Risk and control assessments maintained for CPS 230 purposes may already capture the relevant operational dependencies. An entity with a CPS 230 material service provider register documenting an AI-embedded arrangement may already have partial compliance infrastructure for APP 1. The Guidance should make that connection visible - both to reduce duplication and to help regulated entities recognise where their existing governance documentation already reaches the APP 1 obligation.

A RACI matrix, a contract schedule or a vendor responsibility matrix may help evidence how accountability is distributed across a reliance chain. None of those artefacts, by itself, demonstrates that accountability survived the delegation. The test is whether there is an accountable locus - inside the entity - that can stand behind reliance on the decision or action as a whole, and account for consequences when they emerge.

Question 10 - Extent of disclosure

Governance in Effect as the disclosure adequacy standard

The OAIC's five criteria for effective and meaningful disclosures - plain language, structured for further inquiry, appropriately tailored, logically organised and framed to allow challenge - are sound principles. This submission proposes one additional criterion, which is intended to be distinctive from the existing five rather than overlapping with them.

Disclosures about ADM should reflect **Governance in Effect** - not only governance intent.

A disclosure can satisfy all five of the OAIC's criteria and still describe only what the entity intended. It can identify the type of AI used, the category of decision made, the categories of personal information involved, the policy framework in place and the approval process followed. None of those disclosures, individually or collectively, tells an individual - or a regulator - whether governance actually operated when their information was used to make a decision about them.

Explanation is not the same as warrant

A system-generated rationale, confidence score, trace or summary may assist explanation. It does not determine whether organisational reliance was justified. Explanation matters because it is part of accountability. Warranted Reliance requires more: a demonstrable basis on which the organisation can say why it relied on the AI-enabled decision, why that reliance was within bounds and why it can stand behind the result.

This distinction has direct consequences for disclosure design. An entity that discloses the existence of an AI system and its general purpose has provided an explanation. It has not shown that governance was in effect when the system acted. The Guidance should reflect this gap - not by requiring disclosure of operationally sensitive detail, but by directing entities to address whether governance operated in practice, not only whether it was intended.

Demonstrability, not documentation

Regulatory expectations are converging on demonstrability rather than documentation. Governance must be evidenced through the life of a system, not asserted at the point of approval. A disclosure that accurately describes a governance structure approved at the time of system deployment does not demonstrate that the structure operated - or remained operative - when an individual's information was used to shape a decision affecting their rights or interests.

This is particularly relevant for APRA-regulated entities, which already maintain substantial governance documentation under CPS 230 and related prudential standards: material service provider registers, risk and control assessments, operational incident records and breach documentation. An APP 1 ADM disclosure for a system covered by those frameworks should be coherent with those governance artefacts - not a separate document layer constructed on top of them. A disclosure that accurately describes how an AI-embedded arrangement was governed at the time of system onboarding, but does not reflect whether that governance remained operative as the system evolved, represents governance intent. The individual reading the privacy policy has no basis to know whether the disclosure reflects what the entity approved in an earlier period or what actually governs the system today.

For regulated entities subject to both frameworks, the governance-in-effect standard has a practical implication: disclosures should be updated when the underlying governance arrangements change - not only when the system itself changes. A system that was governed one way at procurement and a different way in current operation requires a disclosure that reflects the current operating reality, not the initial approval.

Chain-of-reasoning traceability

The disclosure should indicate how the entity ensures that the path from trigger to decision - the inputs used, the context applied, the output generated, the human acceptance or modification, and any downstream action - is captured and available for review. This does not require disclosure of commercially sensitive operational detail. It requires the entity to indicate that the reasoning pathway is traceable: that the organisation can show what happened, not merely that a system was used.

Warranted AI's second condition, Chain-of-Reasoning Auditability, defines traceability as contemporaneous evidence that traces the material path from trigger to output and from output to action. A disclosure that describes the type of AI without indicating whether this pathway is documented falls short of the governance-in-effect standard. The individual reading the disclosure, and the regulator reviewing compliance, should be able to identify not just that AI was used, but that the entity can account for how it operated in the specific decision.

Supervisory capacity

The disclosure should indicate what oversight the entity maintains over the ADM system, and whether that oversight is positioned to affect reliance - not merely to record it. A statement that human decision-makers are involved does not suffice; nor does the existence of a monitoring capability that depends on the system's own outputs. The Guidance should encourage entities to describe their oversight posture in terms that distinguish substantive supervisory capacity from nominal human presence.

Together, these elements shift the disclosure standard from governance intent - 'we approved this system, it sits within policy, humans are involved' - toward governance in effect: 'we can trace what happened, we maintained oversight capable of affecting reliance, our disclosures reflect the system as it currently operates and we can stand behind the consequences if tested.' Entities that build their disclosures toward that standard will be better placed when regulatory, legal or affected-party challenge comes.

[redacted]
Governance and risk adviser
[redacted]