Attune Legal

hello@attune.legal
+614 4978 0332

ABN 66 146 013 588
360 Bell St,
Preston VIC 3072

Attune.Legal

Submission to the Office of the Australian Information Commissioner

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

Submitter: [redacted] Attune Legal

Date: 22 May 2026

Contact: [redacted]

Preliminary Observations

Five principled observations underpin the responses that follow:

1. Decision-shaping increasingly occurs upstream of the formal decision point. Automated systems perform an expanding share of their decision-shaping work through intake screening, triage, prioritisation, summarisation, and routing, all of which operate before a human formally makes a decision. Guidance that confines the obligation to the formal decision point will systematically under-capture the activity Parliament intended to address.

2. Human decision-makers cannot detect or override what they cannot see. Filtering performed before a file reaches a decision-maker shapes the

decision more comprehensively than the human's final approval. "Human in the loop" architecture, taken on its own, is not a sufficient indicator that a system is not substantially related to the decision.

3. Workflow-level prospective disclosure is not achievable. Institutions of any meaningful size cannot prospectively map their own systems, because the systems change continuously, overlap in their handling of personal information, and combine in ways that emerge from individual workflow configuration rather than centralised design. The right response is not more granular disclosure, but inspectable retainable configuration of the systems that perform decision-shaping, supported by categorical disclosure in the privacy policy and operationally available contestability rights.

4. Automated systems shape decisions but cannot bear responsibility for them. Responsibility, in the legal sense, can attach only to human or corporate actors. Where the configuration of a decision pipeline routes the visible accountability to a human decision-maker at the end of the pipeline, while leaving the upstream systems and the entity that procured, deployed, and configured them at one or more removes from individuated accountability, the result is that responsibility is concentrated on the actor with the least influence over the substantive content of the decision.

5. Decision-shaping in modern enterprise software is the emergent product of diffuse contributions. Modern enterprise software, particularly software incorporating AI components, is procured for a range of purposes over multi-year horizons, configured incrementally by many staff operating under different functional KPIs, and produces decision-shaping outputs that emerge from the interaction of its components rather than from any unified intent at any single node. "Arranged for" cannot coherently be read to require locatable individual intent in respect of the specific decision-shaping that occurred, because such intent does not, as a matter of organisational reality, exist anywhere in the

institution. The obligation must therefore attach at the level of the legal entity operating the pipeline.

These observations inform each response below.

Question 1 — Substantially and directly related to making a decision

OAIC question:

> *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: Degree of reliance on the ADM system output; Ability and likelihood of human override over an ADM decision; Nature of the output (advisory vs determinative); Transparency and explainability of outputs; Integration of ADM into decision-making workflow; Other. Please rank the factors using the (1-5) ranking system and include any additional factors.*

Ranking of the listed factors:

- Degree of reliance on the ADM system output: **5**
- Ability and likelihood of human override over an ADM decision: **3**
- Nature of the output (advisory vs determinative): **4**
- Transparency and explainability of outputs: **3**
- Integration of ADM into decision-making workflow: **5**

Free text comments:

The two most important factors are **degree of reliance** and **integration into workflow**, because these track the system's actual influence rather than its formal architecture. The factor most likely to be over-weighted in practice is **ability of human override**: override capacity is routinely present in modern systems but rarely exercised, particularly under time pressure or where the human must document reasons for divergence. Formal override capacity should not be treated as evidence of non-substantial relation.

A functional rather than formal test is also necessary to prevent the structure of human-in-the-loop architecture from operating as an accountability sink, in which an automated pipeline's outputs are ratified by a human at the end of the chain whose only realistic role is to apply the system's recommendation. Where the formal architecture concentrates visible accountability on the human, while the substantive decision-shaping has been performed by the system, formal compliance with the obligation can be achieved without giving effect to its purpose. The functional test prevents that outcome.

Additional factors the OAIC should expressly recognise:

- **Order of presentation.** Whether the system's output reaches the human before or after the human forms an independent view. Pre-formed outputs anchor decisions; post-hoc outputs check them.
- **Base-rate concordance.** The empirical proportion of cases in which the human decision matches the system output. Concordance above a threshold should raise a rebuttable presumption of substantial relation, regardless of formal "human in the loop" architecture.
- **Cognitive load and time pressure.** Where decision-makers operate under volume or time constraints, reliance on system outputs is materially higher and "oversight" becomes nominal.
- **Cost asymmetry of disagreement.** Where overriding requires the human to document reasons, escalate, or accept personal accountability for divergence, the system is functionally determinative even where formally advisory.
- **Position in the decision pipeline.** Systems performing intake filtering, triage, prioritisation, or summarisation upstream of a

formal decision-maker are substantially related to the eventual decision, because they determine which matters reach a decision-maker, on what evidence, and with what framing.

- **Information curation.** Whether the system selects, omits, or characterises the information on which the human bases their decision. Curation of the information environment is decision-shaping in substance.
- **Retrospective indistinguishability.** Where a decision-maker cannot, after the fact, identify whether their decision was formed on the underlying record or on an automated summary or characterisation of it, the system should be presumed to have substantially shaped the decision. AI-generated email summaries in widely-deployed productivity tools — such as Microsoft Copilot summaries of email threads — are a paradigm case: a decision-maker who reads a Copilot-generated summary before, or instead of, the underlying correspondence may subsequently be unable to identify whether their view was independently formed or was anchored by the characterisation in the summary. The retrospective inability to isolate the system's contribution is not a reason to treat the system as outside the obligation. It is an aggravating feature: decision-shaping that cannot be traced after the fact is decision-shaping that cannot be contested after the fact.

A principled summary: the obligation should be calibrated to the **functional** influence of the system, not its **formal** architecture. The same software can be substantially related in one configuration and not in another, depending on workflow, presentation, and override behaviour.

Question 2 — Substantially and directly related (NACIA / GPTea Edge Case)

OAIC question:

> *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. Would you consider NACIA's described uses of GPTea to be substantially and directly related to making decisions? (Yes/No) Can you provide your reasons?

Yes.

Reasons:

- GPTea performs two distinct functions, each of which materially shapes the eventual decision. It **summarises** the candidate profile (controlling what the human sees and how it is framed), and it **issues a recommendation** (anchoring the human's evaluation before independent analysis is possible).
- A summary is not a neutral compression. It selects, omits, and characterises. Where the human's decision is made on the summary rather than on independent reading of the underlying record, the summarisation is decision-shaping in substance, not merely advisory.
- The fact that NACIA staff "always make the final decision" is a procedural fact, not evidence of substantive independence. Where the human's information environment is curated by the system, and the system's recommendation is presented before the human forms an independent view, the system substantially facilitates and is directly connected to the decision within the meaning of the obligation.
- This use case should be treated as a paradigm of why human-in-the-loop architecture, on its own, does not take a system outside the obligation. To hold otherwise would convert the obligation into a procedural formality that any entity could discharge by inserting a human ratifier at the end of an automated pipeline.

The principled point is that the human at the decision point cannot override what was already filtered or framed before they saw the file. The aged-care eligibility assessment is therefore made, in substantive

terms, by the chatbot and the human together, and the obligation should attach accordingly.

Question 3 — Significantly affect rights or interests, factors

OAIC question:

> *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. Please rank the examples using the (1-5) ranking system and include any additional examples.*

Ranking of the listed factors:

- Sensitive information: **5**
- Vulnerable persons: **5**
- Intrusive practices: **4**
- Financial outcomes: **4**

Free text comments:

Additional factors the OAIC should add:

- **Reversibility of the decision.** Decisions creating durable records (credit refusals, employment screening outcomes, insurance declinations) have prolonged downstream effects even where any single instance seems minor.
- **Aggregation across decisions.** Individually trivial decisions (price points, ad targeting, content ranking) can, in aggregate, materially shape an individual's opportunity set, financial position, or

information environment. A "course of conduct" lens is more accurate than a per-decision lens.

- **Asymmetry of awareness.** The individual's inability to detect that a decision was made (ads not received, prices not seen, content not surfaced, matters filtered out before review) is an aggravating factor, not a mitigating one, because it forecloses self-help.
- **Suppression and non-escalation.** Decisions that take the form of "failing to do a thing" (failing to escalate a matter for human review, failing to surface a candidate to a recruiter, failing to display an opportunity to a user) should be expressly recognised, consistent with the statutory definition of decision in APP 1.9(b).
- **Effect on the individual's ability to exercise other rights.** ADM that shapes whether an individual can access review, complaint, or appeal mechanisms (for example, by routing them to automated channels that loop without resolution) should weigh heavily, because it compounds opacity with disempowerment.

Question 4 — Vulnerable classes

OAIC question:

> *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?*

Vulnerability should be treated as **contextual rather than categorical**. It is a function of the interaction between the person, the decision context, and the system, not a fixed property of the person.

Classes that should be expressly recognised in guidance:

- Children and young people, including those approaching but below the age of majority;
- People in financial hardship, including those with thin credit files or non-standard income (freelancers, gig workers, casual workers, recently arrived migrants);

- People with cognitive disability, acquired brain injury, neurodivergence, or mental illness;
- People with limited English proficiency or low digital literacy;
- First Nations peoples, particularly in contexts where ADM systems are trained on data that under-represents or mis-represents them;
- People in coercive or controlling relationships, where ADM around finance, housing, telecommunications, or insurance can compound abuse;
- People in acute states (grief, illness, family crisis, displacement) where decision-making capacity is temporarily reduced;
- Older people, particularly those navigating aged-care, health, and financial systems while managing cognitive change;
- People with insecure housing, including those in temporary accommodation or rough sleeping;
- People affected by prior automated decisions (for example, a credit refusal generated by ADM that then conditions further automated assessments), creating compounding vulnerability across systems;

The principled basis: vulnerability for ADM purposes is best understood as **reduced capacity, in the relevant decision context, to detect, contest, or recover from an automated decision that goes against the individual's interests**. Guidance framed around this functional definition will capture the categories listed above and will adapt as new contexts emerge, rather than ossifying around a closed list.

Question 5 — Significant service or support

OAIC question:

> *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. Please rank using the (1-5) system.

Ranking of the listed factors:

- Physical assistance/support: **5**
- Financial assistance: **5**
- Access to educational services: **4**
- "In home services": **4**
- Access to essential banking and credit services: **5**
- Access to telecommunications: **5**
- Access to essential utilities: **5**

Additional categories the OAIC should recognise:

- Healthcare services, including triage, prioritisation, and access to specialist care (consistent with paragraph 343 of the Explanatory Memorandum)
- Insurance (health, home, contents, life, income protection, travel)
- Housing, including rental application screening, social housing allocation, and tenancy databases
- Employment screening, recruitment shortlisting, and workplace allocation systems
- Government benefits, concessions, and entitlements
- Identity verification services, where failure cascades across multiple downstream services
- Legal services and access to justice (for example, automated triage in legal aid intake)
- Child and family services, where ADM informs prioritisation of investigations or interventions
- Disability and aged-care assessment and service allocation
- Educational admissions, scholarships, and placement decisions

Free text comments:

Significance should be assessed by the **role of the service in the individual's life**, not by the dollar value of any single transaction. Telecommunications, for example, are not a discretionary purchase; they are the infrastructure through which an individual accesses employment, healthcare, banking, and government. ADM that affects access to telecommunications therefore affects the individual's access to almost every other significant service. The guidance should treat services with this character (infrastructure-like, gateway services) as significant by default.

Question 6 — Other legal frameworks for "rights" and "interests"

OAIC question:

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

The OAIC should have regard to:

- The International Covenant on Civil and Political Rights and the International Covenant on Economic, Social and Cultural Rights, to which Australia is a party
- The Convention on the Rights of the Child, in particular the best-interests standard
- The UN Convention on the Rights of Persons with Disabilities
- The Disability Discrimination Act 1992 (Cth), the Sex Discrimination Act 1984 (Cth), the Racial Discrimination Act 1975 (Cth), and the Age Discrimination Act 2004 (Cth)

- The Australian Human Rights Commission Act 1986 (Cth)
- State and Territory human rights instruments, including the Charter of Human Rights and Responsibilities Act 2006 (Vic), the Human Rights Act 2004 (ACT), and the Human Rights Act 2019 (Qld)
- The Australian Consumer Law, particularly the prohibitions on unconscionable conduct, misleading and deceptive conduct, and unfair contract terms
- The National Consumer Credit Protection Act 2009 (Cth) and the Banking Code of Practice
- The Fair Work Act 2009 (Cth) in relation to employment-related ADM
- The Freedom of Information Act 1982 (Cth) and equivalent State/Territory regimes, particularly the operational information / Information Publication Scheme provisions
- The Administrative Decisions (Judicial Review) Act 1977 (Cth) and common-law judicial review principles, including the right to a statement of reasons
- The Australian Government's AI Ethics Principles and the Voluntary AI Safety Standard
- Emerging international instruments, including the EU AI Act, the Council of Europe Framework Convention on AI, and the OECD AI Principles, as comparative reference points

---

Question 7 — Differential pricing Edge Case (Daintree)

OAIC question:

> *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 (the luxurious Capital Crescent in Sydney CBD). His desired book, The Art of Inheritance, was priced at $100, while his book-club buddy Charles, from a country town, paid only $35. Marco reads that Daintree*

> *uses differential pricing to inflate prices for individuals from wealthy postcodes. Marco is a freelance artist and cannot afford the book. Restricting access to financial products or healthcare would clearly affect interests significantly. In this Edge Case, would you consider the inflation of prices to have significantly affected Marco's interests? If yes, at what amount would differential pricing be considered significant (5%, 20%, or other)? Rank, on a scale of 1 to 5, (a) the dollar amount of differential pricing and (b) the significance of the product or service.*

Has Marco's interest been significantly affected: Yes.

Reasons:

- Differential pricing based on personal information (postcode used as a proxy for wealth) imposes a financial cost on the individual without their knowledge or consent, and forecloses the ability to comparison-shop on equal terms. This is a substantive interference with the individual's economic decision-making, not merely a market preference.
- Postcode-based pricing operates as a proxy for socioeconomic status and, depending on the area, may correlate with attributes protected under anti-discrimination law. Even where it does not, it produces systematic distributional effects that the privacy framework should not treat as invisible.
- Marco's inability to detect the differential at point of sale (he only learns of it through external reporting) is an aggravating factor, because it forecloses self-help. ADM that produces effects the individual cannot perceive warrants more transparency, not less. If anything, this type of ADM should be prohibited, because it is discriminatory by its very nature – regardless on what basis it determines a buyer’s ability and willingness to pay.

At what amount differential pricing is significant:

The OAIC should resist setting a bright-line numeric threshold (5 percent, 20 percent, or otherwise) because:

- Small percentages on essential or recurring purchases compound into significant lifetime cost

- The same percentage produces materially different effects across income levels
- Numeric thresholds invite gaming (pricing just below the threshold) and create cliff-edges that are difficult to administer
- Significance is properly assessed by reference to the role of the product/service in the individual's life and to the basis of the differential, not by the size of the differential alone

A principled formulation is that differential pricing significantly affects an individual's interests where it is **material relative to the price of the product, applied systematically across a course of dealing, based on personal information the individual did not knowingly provide for that purpose, and not transparent to the individual at point of sale**.

Ranking of the listed factors (1, not important; 5, very important):

- The dollar amount of differential pricing: **3**
- The significance of the product or service being purchased: **5**

Additional factors the OAIC should weigh:

- Whether the pricing variable is a proxy for a protected attribute or socioeconomic status
- Whether the differential is transparent to the consumer at point of sale
- Whether the differential is applied systematically across a course of dealing
- Cumulative effect across the entity's product range
- Whether the individual has practical alternatives at non-differential prices

---

Question 8 — Discriminatory targeted job ad Edge Case (Manscon)

OAIC question:

> *Fictional Edge Case — Discriminatory targeted job ad: Engineering firm Manscon is hiring for graduate engineers. Manscon advertises on job site Opaque Window, which algorithmically promotes the posting to candidates it thinks are best suited. One metric is gender of account holders. Since most engineers at Manscon are male, Opaque Window prioritises promoting the job to male engineering graduates. Rafqa, a female engineering graduate, does not receive the Manscon ad, which limits her employment options. Should it be considered a 'decision' that Rafqa did not receive the Manscon job ad? (Yes/No) Reasons?*

Yes.

Reasons:

- The Privacy Act, on its face, includes "refusing or failing to do a thing" within the statutory definition of decision (APP 1.9(b)). Algorithmic suppression of a job advertisement to a candidate is a "failure to do a thing" (failing to surface the opportunity) generated by the operation of a computer program using personal information (the gender associated with the account).
- Excluding such conduct from the scope of the obligation would render it ineffective against precisely the algorithmic harms most likely to affect employment opportunity, and would conflict with the Explanatory Memorandum's express recognition of effects on access to employment opportunities.
- The fact that Rafqa cannot perceive the decision strengthens, rather than weakens, the case for transparency. Where the individual cannot detect that a decision has been made, transparency through the entity's privacy policy is the only mechanism by which she can become aware of the category of decisions that are being made about her, and exercise downstream rights (including under anti-discrimination law).
- A guidance approach that captures only "decisions" the individual can see (refusals, denials, declinations) would create a structural blind spot in the regulatory regime, because ADM increasingly operates through differential exposure rather than express refusal. Meaning

that the system is biased towards the dominant information within its dataset.

Caveat the OAIC should address in guidance:

The scope of "decision" should not collapse into requiring disclosure of every algorithmic ranking. The principled line is that decisions which **materially affect access to a significant opportunity or service** (employment, credit, housing, healthcare, education) are within scope, even where they manifest as differential exposure or non-escalation rather than express refusal. Guidance can articulate this line by reference to the categories of significance discussed in response to Question 5.

Question 9 — Meaning of "arranged for"

OAIC question:

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

Several scenarios require express guidance, and a structural point underlies them all.

Specific scenarios requiring guidance:

1. Layered procurement. Where Entity A procures a SaaS product from Entity B, which incorporates a model from Entity C, fine-tuned on data brokered by Entity D. Guidance should clarify that "arranging for" follows the entity that determines the **purpose** for which the ADM is used in respect of an individual, regardless of how many vendors sit upstream.

2. Generative AI used by employees on their own initiative. Where an entity tolerates, or has not prohibited, employee use of generative AI tools to draft decisions, summaries, or correspondence affecting individuals, guidance should treat this as "arranging for" once the entity has constructive knowledge. Constructive knowledge should be presumed where the entity has not implemented and audited a policy on permitted AI use.

3. Default-on functionality in standard enterprise software. Where ADM features are enabled by default in widely-deployed productivity suites (for example, AI-generated meeting summaries used in HR investigations, automated triage in case-management systems, AI-drafted assessments in customer service software), the deploying entity has "arranged for" the ADM by failing to disable or configure it. A further and increasingly prevalent instance of this scenario is AI-generated summarisation of communications - for example, Microsoft Copilot summaries of email threads presented to staff before, or instead of, the underlying correspondence. CoPilot is embedded in all Microsoft 365 tools as standard. Google have equivalent tools. Where a decision-maker forms or confirms a view on the basis of a Copilot-generated summary rather than the email itself, the entity has arranged for the decision-shaping that summary performs, by deploying the productivity suite in which the feature is enabled. This is so regardless of whether any individual within the entity consciously adopted the summarisation function for that purpose, and regardless of whether the decision-maker is aware that their information environment was curated by the tool rather than by their own reading of the record.

4. Internal "shadow" deployment. Where parts of an organisation deploy ADM without formal procurement (for example, a team running open-source models locally), the obligation should attach to the legal entity, not to the procurement department. Lack of central visibility is an institutional failure that the entity has accepted by its governance choices, not an excuse for non-compliance.

5. Joint controllers and co-arrangers. Where two entities each contribute to arranging the ADM (for example, a platform and an advertiser jointly configuring targeting parameters), both should be treated as having arranged for the program. Guidance should address how disclosure obligations are allocated between joint arrangers, with the default being that each is responsible for disclosure of its contribution.

6. Configuration-level "arranging". Where agentic AI tools are configured through user-level instruction artefacts (Skills.md files and analogous instruction files in modern agentic systems), the act of writing or deploying the instruction artefact is an act of "arranging for" ADM. The entity that controls the workspace in which such files reside has the means to identify, audit, and govern those artefacts, and should be treated as having arranged for the ADM they configure.

7. Upstream filtering arranged by intermediaries. Where intake, triage, summarisation, or routing systems filter information before it reaches a human decision-maker, the entity using those systems has arranged for the filtering, even if the filtering is performed by a third-party service. Guidance should clarify that "arranged for" reaches the chain of automated steps that shape decisions, not only the final-stage tool.

The structural problem underlying all of these scenarios — diffusion of responsibility:

A further matter requires explicit guidance on the meaning of "arranged for". As a practical matter, modern enterprise software, and particularly software incorporating AI components, is procured for a range of purposes over multi-year horizons, is configured incrementally by many different staff under different functional KPIs, and produces decision-shaping outputs that emerge from the integrated operation of its components rather than from any unified intent at any single node. The procurement decision is not decision-specific, the configuration decisions are layer-specific and locally scoped, and the aggregate decision-shaping behaviour is the emergent product of the interaction of many contributions, none of which was directed at the decision-shaping that ultimately occurred.

If "arranged for" is interpreted to require specific intent in relation to the particular decision-shaping that occurred, the obligation will fail to attach in almost every substantial enterprise context, not because entities are evading it but because the specific intent such a test demands does not, as a matter of organisational reality, exist anywhere in the institution. The same problem produces the diffusion-of-responsibility phenomenon that, more broadly, makes it difficult to ascribe responsibility for upstream processes once filtering by AI components is introduced into a decision pipeline. By the time a decision point is reached, decision-shaping has been performed diffusely across many systems, controlled by many people operating under different functions and KPIs, none of whom can coherently be said to have arranged for the specific shaping that emerged.

The principled response is to read "arranged for" as attaching at the level of the **legal entity that operates the pipeline**, by reference to the entity's structural choice to deploy and run the systems that produced the decision-shaping, rather than by reference to any specific intent located at an individual node. This is consistent with how legal systems already attribute responsibility for emergent organisational outputs in other domains (corporate safety, trust administration, controllership of personal information), and it is the only reading of "arranged for" that produces a coherent allocation of responsibility in the technological environment to which the obligation must apply.

A consequence of this reading is that the absence of identifiable individual intent within the entity is not, and should not be treated as, an answer to a question about whether the entity has arranged for the ADM. Where the entity operates the pipeline, deploys the systems, and benefits from their outputs, it has arranged for what the pipeline does, irrespective of whether any individual within it can be shown to have intended the specific decision-shaping that occurred. Guidance should make this attribution explicit, both to ensure that the obligation attaches in modern enterprise contexts and to prevent the diffusion of

responsibility within institutions from operating as a structural defence against the obligation.

Question 10 — Extent of disclosure

OAIC question:

> *When making disclosures about ADM, the OAIC's view is that entities must ensure that meaningful and accessible information is provided to consumers. The OAIC considers that entities should strike an appropriate balance to ensure disclosures are: clearly articulated in plain language and easy to understand; structured to enable consumers to request further information; appropriately tailored — sufficiently specific to be meaningful, while avoiding overwhelming detail; organised so that similar information is grouped logically; and framed in a way that allows the information and the decision to be challenged or contested. Having regard to these considerations, is there anything additional that entities should disclose to provide effective and meaningful disclosures about ADM?*

The five criteria the OAIC has identified are sound and should be retained. This submission proposes the following additions, refinements, and a structural recommendations.

Additional matters that should be disclosed:

- **The kinds of decisions the individual cannot see.** Where ADM operates through suppression, ranking, differential exposure, or non-escalation, privacy policies should expressly identify the kinds of decisions made by non-action, not only those made by action. This addresses the structural blind spot identified in response to Question 8.

- **The categories of upstream decision-shaping.** Privacy policies should describe, at category level, the kinds of automated steps that shape decisions about the individual, including intake screening, triage, summarisation, prioritisation, and routing, not only the final-stage decision tool. This addresses the structural problem that an increasing share of decision-shaping work occurs upstream of the visible decision point.
- **Effect, not architecture.** Disclosure should describe what kinds of outcomes the system materially affects (price, eligibility, prioritisation, content exposure, suppression), in language a layperson can act on. Telling individuals that "AI is used to assist staff" is insufficient.
- **Pathways to contestability, directed to the entity.** Privacy policies should make clear how the individual can request human review, what evidence to bring, what the entity will do with that evidence, and the timeframe for response. The contact point should identify the **entity**, not the individual decision-maker. The individual decision-maker is not the appropriate respondent to a contestability claim about an entity-level system.
- **Named accountability.** A specific role or contact point within the entity, accountable for ADM transparency, should be identified (likely in most cases to the privacy officer).
- **Layered disclosure.** A short headline notice with progressive disclosure of detail (consistent with the layered approach referenced in international guidance) is preferable to a single block of text.
- **Point-of-decision notification (best practice).** For ADM that materially affects rights or interests, contemporaneous notification at the moment the individual interacts with the system should be encouraged as best practice in addition to the privacy-policy disclosure required by the statute.

A constraint the OAIC should expressly recognise:

Guidance should **not** require entities to publish system inventories, architecture diagrams, or workflow maps. Such disclosure is not achievable for institutions of any meaningful size, for reasons that are structural rather than reflecting bad faith:

- Systems change continuously (patches, vendor changes, reconfigurations measured in weeks)
- Multiple systems process the same personal information for different purposes, with no canonical "the workflow"

- Institutions do not, and cannot, fully know their own systems (shadow IT, unsanctioned SaaS, individually-configured tooling, vendor opacity)
- The composition of effects is emergent from how individual staff configure their workflows, not algorithmic in any documented sense

A disclosure regime that requires the unknowable will produce one of three failure modes: pro-forma boilerplate, immediately-stale honest documentation, or compliance theatre. None of these advances the wellbeing of individuals.

The principled approach is to keep disclosure pitched at the level the statute itself adopts, namely the **kinds of decisions** automated processes shape and the **kinds of personal information** they use. This is stable across system change, meaningful to the individual, and tractable for the entity.

Bridging categorical disclosure and meaningful contestability — configuration auditability:

The gap between categorical disclosure (which is all the privacy policy can practically carry) and meaningful contestability (which is what individuals actually need) can be bridged by **configuration retention and auditability** of the systems that perform decision-shaping. As agentic AI deployment becomes more prevalent, the operative configuration of such systems is increasingly held in primitive, inspectable instruction files (for example, Skills.md files in current agentic platforms, and analogous instruction artefacts in other systems). These files are text, are located in known places in user workspaces, and can be enumerated, version-controlled, collected, and audited because they are written in plain English.

This means that, even where workflow-level prospective disclosure is impossible, an institution **can** in principle reconstruct, after the fact, what filtering and routing was in place at the time of a particular decision, by reference to the configuration files in force at that time. This converts categorical disclosure into something with operational teeth, by ensuring that behind the categorical disclosure there is a

recoverable provenance trail invocable through contestability. Configuration retention is also the mechanism through which entity-level accountability becomes operational rather than nominal: an entity that retains the configuration its systems operated under can answer for what those systems did, while an entity that does not cannot.

The OAIC's guidance under the present Act cannot create a positive obligation to retain such configuration. Guidance can, however:

1. Recognise that, where ADM systems are configured through inspectable instruction artefacts, those artefacts constitute a record of the entity's ADM arrangements
2. Treat retention and version control of such artefacts as good practice, and as relevant to whether the entity has discharged its responsibilities in arranging for ADM
3. Treat absence of inspectable configuration (for example, decision-shaping logic buried in opaque vendor systems with no auditable instruction layer) as a risk factor that the entity has accepted by procurement choice, and that affects the entity's ability to discharge its disclosure and related obligations honestly
4. Encourage standardisation by referring to emerging conventions (Skills.md and analogous primitives) without prescribing any particular instance.

It should be noted that due to the newness of this technology, the platforms currently deploying agentic AI systems with instruction artefacts such as Skills.md files do not yet offer native version control over those artefacts. This is a material limitation: version control of instruction artefacts is a precondition for the configuration auditability pathway described above to function reliably in practice. Without it, an entity cannot, after the fact, reconstruct which version of a Skills.md file was operative at the time of a particular decision. This is because the file may have been rewritten progressively and automatically by the AI system itself in response to subsequent user prompts. A user may find it necessary to refine an instruction artefact mid-conversation, and the AI platform will update the file immediately in response to that prompt. The version state of the configuration and the

history of decisions made under it therefore diverge continuously unless version control is actively maintained.

A positive obligation on individual users or deploying entities to manually maintain version histories of their instruction artefacts is unrealistic and impractical. The purpose of these systems is to reduce cognitive burden on the humans using them. Requiring those humans to maintain parallel version-control discipline as a condition of regulatory compliance undermines that purpose and will not be observed at scale. Human error is an inherent feature of the environment these tools are designed to address; it is not an appropriate foundation for an auditing obligation.

There is no technical reason in principle why version control cannot be built directly into the platforms that deploy and rewrite these files. Platform providers (including Anthropic and its competitors) already manage the infrastructure through which instruction artefacts are created, stored, and updated. The capability required to automatically version-stamp each state of an instruction artefact and retain a timestamped history of each iteration is a standard software engineering function well within the competence of these providers.

This submission respectfully proposes that this is an area in which the law can take a proactive and forward-looking position. A legislative obligation requiring platform providers of agentic AI systems to implement automatic version control over user-deployed instruction artefacts, such that each iteration is retained with a timestamp and can be associated with the decisions made during the period in which that version was operative,would be achievable, proportionate, and technically straightforward for the providers to implement. It would not impose a material burden on those providers, who already manage the relevant infrastructure. It would, however, substantially advance the accountability framework by ensuring that the provenance trail necessary for contestability exists as an automatic byproduct of operation, rather than depending on governance discipline that cannot reliably be

maintained at the user level. The obligation should be stated in technology-neutral terms, by reference to the functional property required, automatic retention of each operative version of any instruction artefact used to configure a decision-shaping system, with timestamps sufficient to associate each version with the decisions made during its operative period, rather than by reference to any particular file format or platform.

A broader recommendation to government, on the public record:

The deeper response to the structural transparency problem is not regulatory adjustment within the existing framework, but recognition that the framework lacks a regulatory primitive adequate to the technology now in use. We respectfully submit that the OAIC should record in its guidance that it will recommend to government the following structural reform:

> *Decision-shaping ADM systems used by APP entities in respect of decisions that significantly affect rights or interests should be required to externalise their operative configuration into a primitive, inspectable file type, capable of being collected, retained, version-controlled, and audited. The format itself is a matter for standardisation; the regulatory principle is that the configuration must exist in a form that survives software change, vendor change, and platform change, and that any auditor of ordinary technical competence can interpret. Skills.md and analogous instruction artefacts in agentic AI deployments are the current dominant instance of this pattern, but the regulatory principle should be technology-neutral and stated in terms of the property (inspectability, primitiveness, auditability) rather than the format.*

Such a primitive would do work across the entire regulatory landscape engaged by ADM, including privacy, anti-discrimination, administrative review, freedom of information, consumer protection, and corporations law. It would convert the present situation, in which institutions do not reliably know their own systems and regulators cannot reliably interrogate them, into one in which the configuration is preserved as an ordinary side-effect of operation, and audit becomes a matter of collection rather than reconstruction. It is the natural complement to the technology-neutral definition of "computer program" already adopted in the Privacy Act, and it would make the ADM transparency obligation deliverable in practice as the technology continues to evolve.

A second pathway — mandatory quality-assurance review of filtered cases:

The conceptual basis of this pathway is that AI-filtered content should be treated, by default, as suspect in the same sense that a manufactured batch is treated as suspect until quality-assurance sampling has confirmed conformance. Quality assurance in other regulated contexts does not assume that the automated process is correct; it tests the output of that process against an independent standard. The same logic applies to automated decision pipelines: the fact that a case has been filtered, triaged, summarised, or prioritised by an automated system is not confirmation that the filtering was accurate or fair — it is a precondition for testing whether it was.

The operational form of this pathway would be as follows. A defined proportion of cases processed through each decision-shaping ADM system are selected for mandatory re-review by a human reviewer who approaches the underlying record independently, without first being presented with the automated summary or recommendation. The re-review is documented and compared against the automated output. Systematic divergence between independent human review and automated output triggers a review of the system's configuration, not merely of the individual case. The results of quality-assurance sampling are retained and available to the regulator on request.

This pathway does not require that every automated decision be reviewed — only that the output of automated systems be periodically tested against an independent human standard. The appropriate sampling rate is a matter for guidance, and may vary by the risk level of the decision class. For decisions affecting access to significant services or significantly affecting rights or interests, a higher sampling rate is warranted.

The QA pathway addresses a failure mode that configuration auditability alone cannot cure: a system whose configuration is perfectly preserved and auditable may nonetheless be producing systematically biased or inaccurate outputs. Configuration auditability establishes what the system was instructed to do; quality-assurance re-review establishes whether what it did was right. Both pathways are necessary for a complete accountability regime. Guidance should recognise both and encourage entities to implement them in combination: configuration retention as the basis for provenance and contestability, and quality-assurance re-review as the operational check on accuracy and fairness.

Closing Observations

The value the OAIC seeks to protect, namely individual wellbeing through informed participation in decisions that shape their lives, depends on individuals having access to information they cannot directly access, and which may not exist in a single place, but is the result of a number of processes that simply resulted in the information they eventually can access. Automated systems, by their nature, occupy positions in the decision pipeline where their influence is largely invisible to the people affected and frequently underestimated by the people deploying them.

A further structural feature of decision pipelines warrants the OAIC's express attention. Automated systems shape decisions but cannot bear responsibility for them. Responsibility, in the legal sense, can attach

only to human or corporate actors. Where the configuration of a decision pipeline routes the visible accountability to a human decision-maker at the end of the pipeline, while leaving the upstream systems (and the entity that procured, deployed, and configured them) at one or more removes from individuated accountability, the result is that responsibility is concentrated on the actor with the least influence over the substantive content of the decision. The human at the decision point cannot detect what was filtered out before they saw the file, cannot audit the configuration of upstream systems before making each decision, and cannot, in any operationally meaningful sense, override what they were never shown.

This accountability asymmetry is reinforced by the diffusion-of-responsibility phenomenon discussed in response to Question 9. Modern enterprise software produces decision-shaping outputs through the interaction of many contributions made under different functional KPIs, none of which is directed at the decision-shaping that ultimately emerges. Within the institution, no individual node bears responsibility for the aggregate effect, because no individual node intended it. The combined result of the asymmetry and the diffusion is a structural state in which the most visible human (the decision-maker) is held accountable for outputs whose substantive shape was determined by actors none of whom is individually accountable. Guidance should refuse to allow this configuration to operate as a structural defence against the obligation, by:

- adopting functional rather than formal tests for substantial relation, so that human-in-the-loop architecture cannot operate as an accountability sink;
- reading "arranged for" at the level of the legal entity operating the pipeline, so that the obligation attaches where it can coherently be discharged;
- directing privacy-policy disclosure to identify the entity and the entity-level contact point for contestability, so that affected individuals are routed to the actor that bears the responsibility rather than to the most visible person in the chain; and

- treating configuration retention and inspectable instruction artefacts as the means by which entity-level accountability becomes operational rather than nominal.

The wellbeing the OAIC's framework seeks to protect therefore has two dimensions. The first, and primary, is the wellbeing of individuals whose rights or interests are affected by automated decisions. The second, which is engaged by the same architectural facts, is the wellbeing of human staff placed in roles where they are expected to ratify automated outputs without the practical capacity to audit them. A guidance regime that locates accountability at the entity level, supports it with categorical disclosure, and grounds it in configuration auditability, addresses both dimensions. A regime that allows accountability to default onto the most visible human in the pipeline addresses neither.

We thank the OAIC for the opportunity to make this submission and would welcome the opportunity to engage further on any of the matters raised.







Attune Legal 