[redacted]
[redacted]
[redacted]

23 May 2026

**Office of the Australian Information Commissioner**

Submission to the ADMT Transparency Consultation

**Dear Commissioner**

**PURPOSE OF THIS LETTER**

I write to you in three inseparable capacities. I am a children's rights advocate with a sustained research and practice focus on trust, consent, and stakeholder agency in the governance of systems affecting vulnerable people. I am a member of the Australian Child Rights Taskforce, The International Association for Safe and Ethical AI; IEEE's Society on Social Implications of Technology, and a graduate of Stanford University's Ethics, Technology, and Public Policy for Practitioners program. I am also a non-executive director of a large non-profit public health service, an industry superannuation fund financial services organisation, and an essential public services trust. In my professional and civic roles, I sit precisely where the governance failures this submission describes are both debated and felt.

Personally, I am the mother of three young people aged 16 to 25. My children have been acutely exposed to these issues through three distinct lenses: an extended period of pandemic quarantine in Melbourne, heightened public exposure as members of a national political family, and their place in the earliest target market of the social media industry.

I believe our regulators have a rare opportunity to prevent the measurable harm that social media was primed to cause. Unregulated Automated Decision-Making (ADM) technology will cause even greater harm as it normalises, creating systemic issues that current frameworks are simply not equipped to address. The window for architectural intervention is narrowing. We must act before ADM becomes so deeply embedded in our health and social services systems that reform requires dismantling rather than designing.

**Because we owe to the child the very best we have to give**

Children are among the least visible subjects of automated decision-making in education, health, and social services, yet they are among the most consequentially affected. Because they have no legal standing to contest decisions made about them, they depend entirely on adults; parents, carers, educators, clinicians, and caseworkers to identify errors in their care. In turn, those adults rely on institutions to maintain the governance architecture that makes those errors identifiable in the first place.

An algorithmic system that shapes a child's academic streaming, standardised results, clinical triage priority, placement in a medical trial, eligibility for a disability support package, or risk classification in a child protection context makes decisions that echo across decades. In many cases, these decisions are made before a child has the capacity to

understand what has happened to them, let alone contest it. By the time they reach adulthood, the causal chain between the algorithmic output and the outcome in their life is completely invisible. Like reaching maturity with a "digital rap sheet," their adult life begins with a deficit or a judgment that no previous generation has had to carry.

This model took hold during the 2020 pandemic, when educational platforms effectively tied children to the internet and social media spaces, eroding parental choice and oversight. Children in the populations this submission focuses on most children in poverty, children with disabilities, children from culturally and linguistically diverse (CALD) families, and children whose parents are experiencing family violence or housing insecurity, face this "double exposure" problem in acute form. Their circumstances are those that algorithmic systems are designed to assess, yet their circumstances are also the factors that eliminate the advocates, the literacy, and the institutional access required to contest algorithmic error.

A **child cannot consent to a system they do not know exists**, cannot understand a process no one has explained, and cannot contest a decision when no adult in the governance chain even knows an algorithm was involved. The consent and accountability failures this submission identifies are not abstract; for children, they are developmental. The harm compounds across a lifetime.

Trust in this context is not a vague thing. It is a reasonable expectation, grounded in transparency, accountability, and genuine responsiveness to harm, that the people and institutions making decisions about a child have that child's long-term interests at heart. Automated decision-making, as currently deployed and governed, cannot sustain that expectation. It cannot even be held accountable to it, because no one in the governance chain can say with confidence what the system is doing, or to whom.

**As a board director**

I also write as someone who holds the exact governance responsibilities this submission asks the regulatory body to support. As a board director of a non-profit health and social services organisation, I am accountable for the care we provide to people who, by the very nature of our mission, are among the most vulnerable in our community.

I know from direct experience that the invisibility problem identified in this text is a live risk. Strategic investment decisions in clinical and operational capability routinely include technology platforms whose algorithmic components are not visible in board papers, have not been assured for equity performance, and are not monitored for population-level outcomes after deployment. The board approves the platform, and the algorithm arrives with it.

Without a regulatory framework that creates explicit obligations, I am not in a position to independently identify every ADM system embedded in our operational environment, commission equity impact assessments for each one, establish subgroup performance monitoring, and design consent architectures that satisfy a meaningful comprehension standard.

I have the governance *will* to do these things. What I lack is the regulatory *mandate* that would give me the organisational leverage to require them of vendors, of management, and of the procurement processes through which technology enters our organisation.

That is what this submission asks the regulatory body to provide. It is not a call for new burdens on already-stretched non-profit boards. Rather, it is a plea for a framework that gives boards the mandate, the language, and the regulatory backing to ask the right questions, require the right disclosures, and hold vendors and management to account for the equity consequences of the technology they deploy. Board directors of non-profit organisations want to govern well; **we need a regulatory architecture that makes good governance achievable.**

**Structure of my submission**

This submission is laid out in five parts. I start with the real, often hidden harm that automated decision-making is currently causing our most vulnerable people. From there, I look at how we've started to let "machine authority" become the norm and what that means for our democracy. I then make my argument in what Australian leaders like Edward Santow and Toby Walsh see as essentials, natural justice and human rights, before moving into the "how-to" of change: a three-part reform package centred on the crucial role of board directors of for-purpose organisations.

To make this as practical as possible, I've included two supporting documents:

- **Annex A** sets a clear "comprehension standard" for meaningful consent. It's designed to give the regulator a practical benchmark to see if people actually understand what they are signing up for, particularly in high-risk settings.
- **The Board Director Explanatory Note** is written in plain language for directors who might not have a technical background but need to know which questions to ask at their next board meeting and what governance steps to take now.

This work is built on my original post-graduate research, in the foundation of research from major institutions, from the latest findings in courts, the OECD and UNICEF, to the terrible lessons of the Robodebt Royal Commission. It also draws deeply on my work regarding trust and agency for people affected by these systems. Most importantly, it has been shaped by the lived experiences of children in our public services, whose rights have both informed and tested every recommendation I make.

**What I ask of our public service**

I ask the OAIC to view this submission as the starting point for a binding regulatory standard. The harms documented here are neither potential nor theoretical. As the Robodebt Royal Commission, which I witnessed at close quarters, showed us: **when automated systems are deployed without consent, transparency, and accountability, people get hurt**. This harm builds up quietly in the shadows before it breaks into the open, and it always hits hardest those who have the least power to fight back.

Our children, especially, cannot afford to wait for regulation to catch up, and keep up, with the technologies already shaping their lives and their futures. The governance architecture proposed in this submission, including an Automated Decision-Making (ADM) Register, clear consent and right-of-reply protections, and rights-based stakeholder impact assessments, is the minimum structure that the evidence demands.

I would welcome the opportunity to sit down with the OAIC to discuss any part of this submission. I am very keen to provide further evidence, speak to the critical board-level governance reforms needed, or talk through the children's rights analysis in greater detail.

Yours sincerely,

[redacted]

Children's rights advocate

Board director, AI Governance and Ethics

**Attachments**

1. Governing ADM in health and social services full submission (parts a–e)
2. Annex A: a comprehension standard for meaningful consent to ADM in clinical settings
3. Explanatory note for board directors: automated decision making in health and social services

# Governing Automated Decision-Making Technologies with (not for) vulnerable Australians.

**A submission to the Office of the Australian Information Commissioner consultation on ADM technology transparency guidelines.**

15 June 2026

[redacted]
Sector: Public services and board governance

This submission draws on my research into trust frameworks, consent mechanisms, and stakeholder agency. I argue that automated decision-making (ADM) in service delivery is a profound governance, democratic, and human rights issue requiring regulatory intervention.

## Summary

Automated decision-making (ADM) technology, as currently implemented across health, social services, welfare, and financial systems, constitutes a systemic governance failure with measurable human consequences.
My argument spans three stages:

- **Stage 1: the double exposure problem.** I document the unequal harm experienced by those who are both the primary targets of ADM and the least equipped to fight back. For populations experiencing poverty, housing insecurity, family violence, or disability, their life circumstances are the precise inputs that trigger algorithmic assessment. Yet, these same circumstances eliminate the digital access, literacy, and institutional standing required to challenge a wrong decision.
- **Stage 2: the normalisation of machine authority.** ADM governance is typically siloed as a welfare or consumer protection issue. This is a mistake. What is at stake is the gradual displacement of human deliberation as the legitimate basis for decisions across all domains of civic life. As these opaque systems normalise in public services, the dangerous principle they establish (that an algorithm can make high-impact decisions without transparency or a right of reply) is migrating into credit scoring, employment screening, and even board selection.
- **Stage 3: the rights framework.** I locate this governance failure within three inseparable frameworks: natural justice, international and domestic human rights obligations, and democratic due process. They establish a foundational truth: **the right to be governed by human judgment in high-impact decisions is a democratic requirement and a human rights obligation.**

I propose three structural reforms to establish the regulatory architecture that current ADM deployment has assumed away:

1. A mandatory **board-level governance framework** for ADM oversight, requiring equity reviews as a precondition of deployment.
2. Strengthened **consent, transparency, and right-of-reply obligations** that are genuinely accessible regardless of literacy, language, or digital access.
3. A **rights-based stakeholder impact assessment** co-designed with affected communities before any ADM system enters service delivery.

# Part A: Differential harm and the "double exposure" problem

The literature on ADM harm frequently frames "vulnerability" as a personal deficit, defined as a lack of cognitive, linguistic, or economic capacity to traverse complex systems. In Virginia Eubanks' work on ADM, she argues "vulnerability is a property of the relationship between an individual and a system, not a characteristic of a person."
Her research into "digital poorhouses"automated eligibility systems, predictive policing, and child welfare algorithms shows that vulnerability is manufactured by the design of the system. When built without accountability to the people it serves or targets, it creates a power imbalance in which:
The system contains all the data and decision-making power and the individual has no transparency into why a decision was made and no clear path to appeal it.

## A.1 The double exposure mechanism

What my research regularly finds is a compounding, destructive dynamic called **double exposure**:
Exposure 1: where algorithmic risk has a high risk of error due to unrepresentative training data or overly complicated life circumstances.
Compounds Exposure 2: which is the structural barrier where a person's life circumstances simultaneously strip away the means to identify, contest, or correct the error.
An algorithm does not self-correct when the correction pathway has been structurally removed. A person experiencing housing insecurity may never receive the digital debt notice. A person with a psychosocial disability may lack the help to navigate a rigid online review process. A woman fleeing family violence may have record gaps that a machine flags as fraud.

## A.2 Documented governance failures

This mechanism has affected Australian institutions:

- **The Robodebt Scheme (2016–2019):** Unlawfully issued approximately 470,000 inaccurate debt notices using an unauthorised income-averaging methodology. It represents the ultimate double-exposure failure: recipients were kept in the dark about the algorithmic logic, the burden of proof was reversed, and the review pathway was inaccessible to those in financial hardship.
- **The NDIS planning process:** Algorithmic needs-categorisation consistently fails to accommodate the complexity of psychosocial disabilities. This leaves participants and their families without accessible explanations or review mechanisms.

- **Proxies allowing discrimination in credit scoring:** Financial algorithms routinely use postcodes, employment types, or languages spoken as proxies for race, income, or background, creating systemic bias without explicitly encoding protected attributes.

In all these cases, the technical systems worked exactly as they were designed to. The failure belongs entirely to the governance architecture around them, which did not exist.

# Part B: The normalisation of machine authority

A welfare framing implies that the solution is targeted protection for identifiable vulnerable groups. I propose that unconstrained ADM deployment is quietly normalising machine authority as the default basis for human choices.

## B.1 The reversal of justification

Normalisation does not mean public approval; it means institutional resignation. Repeated exposure trains professionals and the public to accept algorithmic outputs as presumptive authority. In human systems, the burden of justification sits with the decision-maker. When the decision-maker is an algorithm, the burden reverses: the output is presented as objective, and the individual must prove the machine wrong.

## B.2 The invisibility problem in healthcare

In health and social services, ADM rarely arrives labelled as an artificial intelligence risk. It arrives buried in the contract as an enhanced Electronic Medical Record (EMR) feature, a clinical decision support module, or a scheduling optimisation tool. A board approves a strategic technology investment; the algorithm simply tags along.
The real-world dangers of this invisibility are empirically demonstrated by *Stanley, Tsang, Gillett et al. (2026)* in *Nature Communications*. Their 50-year sociotechnical simulation of AI-assisted breast cancer screening revealed a frightening gap between mathematical fairness and human outcomes:

## The clinical fairness issue (Stanley et al., 2026)

Even when algorithms were ‘mathematically optimised’ for "fairness," the mortality rate ratio between privileged and unprivileged groups remained three to four times higher. The system's false-negative rate was consistently higher for the underprivileged group. Crucially, clinicians' reliance on the system amplified this difference through automation bias, leading them to blindly defer to the machine rather than exercising independent judgment.
This variation remained invisible to traditional governance structures. No performance dashboard or board paper captured it. It only became visible when population-level outcomes were tracked over decades. This the kind of long-term monitoring that a governance framework needs.

## B.3 Scope creep and aggregation of power

This structural problem is moving into high-status domains, including automated recruitment screening, board suitability assessments, and corporate creditworthiness evaluations. The lack of transparency and the right of reply are inherent to the technology itself and not just ‘accidental features’ of systems designed for people in poverty. A governance framework that is insufficient to protect a welfare recipient will also fail when applied to a board appointment.

Absent explicit regulatory systems, proprietary tech entities are assuming a social "license to operate" by default, accumulating interpretive authority over human lives without a public or democratic mandate.

# Part C: The rights framework

"The use of AI in decision-making must not lead to a 'black box' where the reasons for a decision are hidden. Human rights must be at the centre of how AI is developed and used, ensuring that technology serves humanity, not the other way around."
**Edward Santow**, Former Australian Human Rights Commissioner

"When data and AI shape public services, children experience their consequences first and longest... Health records, education histories, social protection registries, and child protection case files exist because a child needs support. Children do not consent in the way adults do, and they cannot meaningfully challenge how their data is used. That imbalance is why questions of control, accountability, and governance carry more weight when children are involved."
**UNICEF**, *Policy Guidance on AI and Children*, 2026

"..What makes Australia so special that we'll see the benefits of AI without making the sort of new AI laws that other nations are introducing? ... Social media should have been a wake-up call about the harms of unregulated AI. We're about to supercharge the sort of harms we saw with social media with an even more powerful and persuasive technology."
**Professor Toby Walsh**, National Press Club Address, February 2026

A decision made through a process that cannot be scrutinised, attributed, or corrected is illegitimate. My proposed reforms rests on three  legal pillars:

## C.1 Natural justice and procedural fairness

The common law principles of natural justice, specifically the **hearing rule** and the **bias rule**, are the bedrock of legitimate decision-making in Australian administrative law.

- **The Hearing Rule** is breached because a person cannot respond to a case assembled by a hidden algorithm. A machine does not give reasons; it gives outputs. By the time a patient or client sees the output, the moment for human deliberation has passed.
- **The Bias Rule** is breached because algorithms are not impartial. They encode the priorities and systemic biases of their creators. As *Grgić-Hlača et al. (2018)* demonstrated, people consistently judge the use of features (like postcodes or housing stability) as procedurally unfair, violating basic social intuitions of justice.

In healthcare, this is compounded by the landmark High Court ruling in ***Rogers v Whitaker*** **(1992)**. The Court established that a patient's right to self-determination is paramount, ruling that a doctor has a strict duty to disclose any "material risk" that a reasonable person in the patient's position would consider significant.
When a clinician defers to an invisible algorithm due to automation bias, and the patient is never informed that a machine is screening, triaging, or prioritising their care, the legal standard of informed consent is erased.

## C.2 Human rights frameworks

Australia's obligations under the International Covenant on Civil and Political Rights (ICCPR) and domestic state Human Rights Acts are directly violated by some current ADM practices:

| Right | Algorithmic Breach Mechanism |
|---|---|
| **Equality & Non-Discrimination** | The use of proxy features (income, postcode, language) to produce indirect, discriminatory outcomes against marginalized groups. |
| **The Right to Privacy** | Ingesting health, financial, and behavioural records into training datasets without meaningful, informed consent. |
| **An Effective Remedy** | The complete absence of accessible, human-led contestation pathways to reverse algorithmic errors. |
| **The Right to Participation** | Systematically shutting out patients and vulnerable stakeholders from having a say in how decisions about their lives are shaped. |

### C.3 Democratic due process and the constitutional powers

Democratic governance requires that authority be accountable, attributable, and correctable. The constitutional separation of powers balances public authority across the legislative, executive, and judicial branches.
Unregulated ADM introduces a fourth actor into public services: a proprietary algorithm owned by a private entity that exercises public power. It is not legislatively authorised, not subject to public service accountabilities, and cannot be judicially reviewed because it leaves no human ‘paper trail’.

## Part D: The board director and reforms

### D.1 The Board Director

As a board director of health and social services organisations, I carry fiduciary duties, duties of care, and community trust obligations that current ADM deployment patterns make impossible to satisfy.

- **The Fiduciary Duty** requires the directors to act in the best interests of those they serve. Approving any IT platform without visibility into its embedded algorithms is a blind delegation of care to an unvetted vendor.
- **The Duty of Care** requires taking reasonable steps to prevent harm. The current procurement landscape creates a **trust cascade failure**, which is an information gap stretching from vendor to board, executive to clinician, and finally clinician to patient. When no single actor has visibility, a duty of care exists in name only.
- **The Community Trust Obligation** dictates that the board carries ultimate moral responsibility for its technology. As *Robodebt* proved, public consciousness no longer accepts the excuse that a harmful algorithm was proprietary or outsourced. The buck stops with us at the board level, and that accountability cannot be contracted away.

### D.2 AI's great potential requires responsible deployment

My submission is not an argument against AI. As *Reddy (2026)* notes in *Learning Health Systems*, global healthcare faces an unprecedented convergence of rising costs and a projected deficit of 18

million healthcare workers by 2030. These shortages hit regional and vulnerable populations hardest.
Responsibly deployed AI has a genuine, transformative capacity to democratise care. We can see this in automated diabetic retinopathy screenings and AI-assisted mammography, which catch what human review misses and bring specialist-level diagnostic accuracy to under-resourced communities.
But *Reddy (2026)* provides the conditions this potential needs:
Healthcare systems must mandate comprehensive algorithmic impact assessments before deploying AI in vulnerable communities, backed by continuous bias-monitoring protocols to ensure that AI trained on one population works equitably in another.

***The proposed reforms are not obstacles to technological innovation; they are the absolute prerequisites for its safe, ethical adoption.***

## D.3 The patient consent architecture (Teach Us Consent V2)

The traditional model of informed clinical consent, built on a single moment of disclosure and choice, is broken by embedded ADM.
An algorithm has already made the triage decision before the patient arrives. It has influenced the screening priority before the appointment is scheduled. It dictates the follow-up interval before discharge. Furthermore, as the *Stanley et al. (2026)* simulation brings to light, an algorithm embedded in a screening pathway shapes a patient's diagnosis, treatment priority, and further care across a timeframe extending years beyond a single clinical encounter.

Trust in AI in healthcare requires patient consent in a new, systemic architecture that looks like this:

A Systemic Public Disclosure = Clear, accessible notification that ADM is used within the care pathway.

B Plain-Language Explanation = Explicit mapping of the categories of decisions influenced by the system.

C The Human Circuit-Breaker = A standing, easily executable individual right to request immediate human review.

D Continuous Population Auditing = Ongoing outcome monitoring to detect hidden disparities before harm compounds over time.

A patient discharged from a health service does not cease to be a person whose life and rights our boards are meant to protect. As directors, we want to govern well. I am asking the OAIC for regulatory backing, a mandate, and clear language to hold vendors to account and make good governance achievable.

# Submission Reform 2: A comprehension standard for meaningful consent to ADM in health settings

*This document proposes a comprehension standard for regulatory purposes: the minimum conditions that must be satisfied for a person's consent to ADM-influenced clinical decision-making to be regarded as meaningful. It is intended for use by the regulatory body in assessing the adequacy of disclosure and consent practices, and by organisations in designing consent architectures that satisfy the rights established in Reform 2 of the main submission.*

## A.1 Why standard clinical consent is insufficient for ADM

Informed consent in clinical practice rests on a well-established doctrine: the patient must receive sufficient information, in a form they can understand, to make a decision that reflects their own values and circumstances. Australian law, following the High Court's decision in Rogers v Whitaker (1992) and the subsequent development of the material risk standard, requires that a clinician disclose any information that a reasonable person in the patient's position would want to know, and any information that this particular patient would want to know.

This standard was developed for human clinical decision-making. Applied to ADM embedded in care pathways, it produces structural comprehension failures that make the standard consent model inadequate as a governance mechanism. This is because the conditions for comprehension do not exist and cannot be created through disclosure at the point of care alone.

## A.2 Comprehension failures

### 1. The opacity failure

Comprehension requires knowing that an algorithm exists and has influenced the decision in question. This is the threshold condition: In current clinical ADM deployments, this condition is frequently not met because the trust cascade analysis demonstrates that the information does not reliably reach the clinical encounter at all. The board approved a platform. The executive manages a system. The clinician uses a tool. The patient interacts with a clinician. At no point in this chain does a disclosure obligation currently require that the algorithmic component of the decision be identified and communicated.

> *A consent framework for ADM must begin with a mandatory identification obligation: every system in clinical use that contains algorithmic decision-making functionality must be identified to the patient as such, in plain language, before or at the point of the clinical encounter it influences.*

### 2. The explainability failure

Comprehension of how an algorithm influenced a decision requires an explanation of the system's logic in terms a patient can evaluate. Many clinical AI systems (particularly those using deep learning architectures) are not interpretable in this sense. Their outputs cannot

be traced to explicit reasoning that a clinician, let alone a patient, can assess. This is not a marginal problem affecting a minority of systems. It is endemic to the deep learning approaches that underpin the most widely deployed clinical AI tools, including some radiology and pathology AI systems.

Grgić-Hlača, Zafar, Gummadi and Weller (2018) establish that procedural fairness ( the dimension of fairness most relevant to consent) requires knowledge of an output but also the ability to assess the legitimacy of the process that produced it. Knowing that 'the system flagged your case' is not comprehension. Knowing on what basis, using what features, with what known limitations for people with your profile is the minimum information on which comprehension could rest.

> *Where a clinical AI system cannot provide an interpretable explanation of its outputs in terms accessible to a non-specialist, this limitation must itself be disclosed. A patient who is told 'the system cannot explain how it reached this assessment' has received information material to their decision about whether to accept or request human review of the algorithmically-influenced recommendation.*

## 3. The specific population failure

Aggregate accuracy statistics are not a sufficient basis for comprehension.
A patient told that an AI system is 94% accurate has been given information that may not describe their risk at all. Stanley, Tsang, Gillett et al. (2026) demonstrate through socio-technical simulation that AI-assisted screening systems can maintain high aggregate accuracy while producing mortality rate ratios of three to four between privileged and unprivileged populations. The accuracy figure tells the underprivileged patient nothing accurate about their risk of being the patient the system misses.

Meaningful comprehension requires subgroup-specific performance data: accuracy, false positive rate, and false negative rate for populations with the patient's relevant demographic and socioeconomic characteristics. This is the information on which a rational decision about accepting or challenging an algorithmically-influenced recommendation could actually be based.

> *Disclosure of an aggregate system performance is insufficient. Organisations must be able to provide subgroup-specific performance data for any clinical AI system in use, and must disclose known performance differentials to patients whose characteristics place them in a group for which the system's performance differs materially from the aggregate.*

## 4. The temporal comprehension failure

Clinical consent is typically structured around a discrete intervention with identifiable proximate risks. ADM embedded in a care pathway shapes decisions across an extended and often indeterminate timeframe: screening priority, triage, treatment sequencing, follow-up intervals, further care referral. The consequences of algorithmic influence accumulate across months and years before they become visible as outcomes. A patient at a single clinical encounter cannot comprehend a risk whose full scope will not be apparent until years later, through a sequence of decisions made by multiple different clinicians relying on a system none of them may know they are using.

Stanley et al.'s fifty-year simulation makes the temporal structure of this harm concrete. The mortality disparity their model identifies does not appear at any single clinical encounter. It emerges from the accumulation of algorithmically-influenced decisions across the full care pathway. No consent obtained at any single point in that pathway encompasses the full scope of what the patient would need to comprehend to give meaningful consent to the algorithm's role in their care.

> *Consent to ADM in clinical settings cannot be a single-encounter event. It must be supported by a standing right to information and review at any point in the care pathway (including after discharge) and by an ongoing organisational obligation to monitor and disclose population-level outcome data that makes the temporal dimension of algorithmic influence visible.*

### 5. A differential literacy failure

Comprehension in clinical consent is already unequal across populations: health literacy, language proficiency, cognitive capacity, and access to support all affect what a person can understand and how they can use that understanding. ADM adds a layer of technical complexity that compounds existing comprehension inequalities in a specific and consequential direction.

The groups with the lowest health literacy and least access to support are precisely the groups for whom ADM systems are most likely to have differential performance characteristics because these groups are underrepresented in training data, because proxy discrimination through features such as postcode and employment type disproportionately affects them, and because the algorithmic systems most widely deployed in health and social services were designed and validated on populations that do not adequately represent them. The person least able to comprehend the system's limitations is the person most likely to be on the wrong side of them.

This means that a uniform disclosure approach (the same information delivered in the same format to all patients) does not satisfy a comprehension standard. It satisfies a notification standard. Comprehension requires that the information be delivered in a form, language, and level of detail appropriate to the individual patient, with access to support where needed.

> *A comprehension standard for ADM consent must be equity-differentiated: the disclosure obligation must be more substantial, not less, for patients in groups known to be at elevated risk of differential algorithmic performance. Interpreter access, advocate support, and accessible plain-language materials are not optional accommodations, they are preconditions of comprehension for the patients most exposed to algorithmic harm.*

## A.3 The comprehension benchmark

The following table sets out the minimum comprehension standard for meaningful consent to ADM in clinical settings. It is proposed for use by the regulatory body as a benchmark against which disclosure and consent practices can be assessed, and by organisations as a framework for designing consent architectures that satisfy the rights established in Reform 2 of the main submission.

| Comprehension requirement | What it demands in practice | Is it currently met? |
|---|---|---|
| Existence of ADM | Patient knows an algorithm was involved in the relevant decision, in plain language, before or at the clinical encounter | Rarely, no mandatory identification obligation exists |
| Nature of influence | Patient understands what category of decision the algorithm influenced (screening priority, triage, treatment sequencing, follow-up interval) | No, typically not disclosed |
| Feature basis | Patient is told what kinds of information the algorithm used (whether volitional or non-volitional features shaped the output) | No, most systems cannot or do not disclose this |
| System limitations | Where the system cannot explain its own reasoning, this limitation is itself disclosed | No, not a current requirement |
| Population-specific performance | Patient receives accuracy, false positive and false negative data for populations with their demographic and socioeconomic profile, not aggregate statistics only | No, aggregate statistics only, if disclosed |
| Temporal scope | Patient understands that algorithmic influence may extend across the full care pathway, not only the current encounter | No, consent is encounter-based |
| Right to human review | Patient is told they can request human review of any ADM-influenced decision, at no cost, at any point including post-discharge | No, no standing right currently exists |
| Accessible support | Interpreter, advocate, or support person access is offered where needed for comprehension, (at no cost to the patient) | Rarely, not systematically provided for ADM contexts |
| Ongoing disclosure | Patient is informed of material changes to system performance or configuration that affect their care pathway | No, no ongoing disclosure obligation exists |

> *On the comprehension benchmark proposed above, current ADM consent practices in health and social services do not satisfy the minimum standard for meaningful consent in any of the dimensions identified. This is not primarily a failure of clinical*

> *communication but a structural failure produced by the absence of the organisational and regulatory obligations that would make comprehension possible.*

## A.4 The systemic consent architecture

Because the five comprehension failures are structural rather than individual, they cannot be addressed through improvements to clinician communication alone. They require a systemic consent architecture; a set of organisational and regulatory obligations that together create the conditions within which individual comprehension can occur.

The systemic consent architecture has four components, each corresponding to one or more of the comprehension failures identified above.

### Component 1. Public disclosure

Every health and social services organisation using ADM in clinical or operational decision-making must publicly disclose that it does so, in a form accessible to current and prospective patients. This disclosure must identify the categories of decision that ADM influences, the population groups subject to those decisions, and the mechanism through which a patient can obtain further information or request human review. This addresses the opacity failure at the organisational level.

### Component 2. Encounter-level disclosure

At or before any clinical encounter involving an ADM-influenced decision, the patient must be told that an algorithm has influenced or will influence the relevant decision, what category of decision is involved, and what their right to human review entails. This disclosure must be in plain language, available in the patient's preferred language, and accompanied by an offer of interpreter or advocate support where the patient's characteristics indicate this is needed. This addresses the opacity failure at the individual level, the differential literacy failure, and the notification precondition for all subsequent comprehension.

### Component 3. On-request explanation

Any patient who asks for an explanation of how an algorithm influenced a specific decision about their care must receive: the category of features the system used; the system's known performance characteristics for populations with the patient's relevant profile; any known limitations on the system's ability to explain its own outputs; and a clear statement of the patient's right to human review. Where the system cannot provide an interpretable explanation, that limitation must be disclosed. This addresses the explainability failure and the population specificity failure.

### Component 4. Ongoing right to review

The right to request human review of an ADM-influenced decision must be a standing right, exercisable at any point in the care pathway including post-discharge, within a timeframe meaningful in the context of the decision, and at no cost to the patient. This addresses the temporal comprehension failure by removing the requirement that comprehension and consent occur at a single point in time, and the differential literacy failure by removing cost as a barrier to exercising the right.

# References

Grgić-Hlača, N., Zafar, M. B., Gummadi, K. P., & Weller, A. (2018). Beyond distributive fairness in algorithmic decision making: Feature selection for procedurally fair learning. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1). https://doi.org/10.1609/aaai.v32i1.11296

Stanley, E. A. M., Tsang, R. Y., Gillett, H., Souza, R., Vigneshwaran, V., Kang, C., McCradden, M. D., Wilms, M., & Forkert, N. D. (2026). Connecting algorithmic fairness and fair outcomes in a sociotechnical simulation case study of AI-assisted healthcare. Nature Communications, 17, 788. https://doi.org/10.1038/s41467-025-67470-5

Reddy, S. (2026). Foundation for artificial intelligence-driven democratization of healthcare access. Learning Health Systems, 10(3), e70093. https://doi.org/10.1002/lrh2.70093

Rogers v Whitaker (1992) 175 CLR 479. High Court of Australia.

Australian Commission on Safety and Quality in Health Care. (2019). Australian Charter of Healthcare Rights (2nd ed.). ACSQHC.

## Explanatory note

# The algorithm comes with the platform: what board directors of health organisations need to know about Automated Decision Making (ADM) technology.

*Prepared for regulatory submission to the OAIC Consultation on ADMT transparency and governance.*

> *This note is written for board directors without a technical background in AI or algorithms. It explains what automated decision-making is, why it is a board governance matter, what your organisation is likely already doing that you may not know about, and what you should do now. It is a companion to a regulatory submission addressed to the sector regulator.*

## 1. What is automated decision making and why does it matter for a board?

Automated decision making, ADM, means any system that uses a computer algorithm to generate an output that influences or determines a decision about a person. It does not have to be labelled 'artificial intelligence' or 'AI' to be ADM. It includes clinical decision support tools that recommend a treatment or flag a case for review; scheduling systems that prioritise appointments; risk-scoring systems that assess a patient's need for follow-up; workforce management tools that allocate staff; and financial assessment tools that determine eligibility for services or supports.

ADM matters for boards because decisions made algorithmically are still decisions the organisation is responsible for. If an algorithm embedded in your electronic medical record system produces a recommendation that contributes to a patient receiving lower-priority care than they needed, the fact that a machine made the recommendation does not discharge your organisation's duty of care. In the eyes of patients, regulators, and the public, your organisation made that decision.

> *You do not need to have approved an AI system for your organisation to be using one. ADM is frequently embedded in platform upgrades, new software modules, and vendor-supplied enhancements to systems your board approved for other purposes. The algorithm was in the contract, not the board paper.*

## 2. What is your organisation probably already doing that you don't know about?

Research into AI deployment in education, healthcare, including a 2026 simulation study showed that AI systems embedded in clinical workflows are frequently invisible to the governance structures responsible for them. This is a structural consequence of how these technologies are procured and deployed.

The organisation is likely already using ADM in one or more of the following ways:

- Clinical decision support: your EMR or radiology system may include AI tools that flag cases for review, suggest diagnoses, or recommend treatment pathways
- Scheduling and triage: patient appointment priority or emergency triage may be influenced by algorithmic risk scoring
- Workforce management: rostering, allocation, and performance evaluation systems may use algorithmic optimisation
- Financial assessment: eligibility assessment for services, supports, or financial assistance may involve automated scoring
- Clinical trials: participant matching or eligibility screening for research may be algorithmically assisted

The question for your board is not whether any of these uses is appropriate. Some are highly beneficial. The question is whether we know which systems contain ADM, what they do, who they affect, how well they perform for different stakeholder groups, and what happens when they get it wrong.

## 3. Why are your most vulnerable patients at greatest risk?

The patients most likely to be affected by ADM-driven decisions, people accessing welfare supports, disability services, mental health care, aged care, emergency housing, and low-cost clinical services, are the same patients whose circumstances are most likely to be flagged and assessed algorithmically. They are also the patients least likely to know that an algorithm shaped a decision about their care, and least equipped to challenge it if something goes wrong.

This matters for two reasons. First, the research evidence shows that AI systems trained on data that underrepresents these populations perform less well for them. This may mean producing more errors, more missed diagnoses, and more inappropriate recommendations for the people who most need accurate care. Second, when an algorithm makes an error for a patient who has limited English, no fixed address, a cognitive impairment, or no advocate, that error is far less likely to be identified and corrected. The harm accumulates silently.

> *If the organisation deploys an AI system that produces better outcomes for patients from higher socioeconomic backgrounds and worse outcomes for patients from lower socioeconomic backgrounds and we do not know this is happening because we have no mechanism to monitor it, your board has not discharged its duty of care, regardless of who supplied the system.*

## 4. Can your patients consent to ADM in their care?

Probably not in any meaningful sense, under current conditions. And this is not primarily about the patients.

Meaningful consent requires, at minimum, that a person knows an algorithm influenced a decision about them, understands what it did and on what basis, and has a genuine option

to request human review. None of these conditions are routinely met in current clinical ADM deployments, for reasons that have nothing to do with patient capacity.

The clinician treating the patient may not know the decision support tool they are relying on is algorithmic. As a board director, we may not know which systems in your organisation contain ADM. Your CEO and clinical leadership may not know either. A patient cannot consent to what no one in the governance chain has identified.

There is also a timing problem. Algorithms embedded in care pathways shape decisions before, during, and after the clinical encounter, screening priority, appointment sequencing, follow-up intervals, treatment pathways. By the time a patient is sitting with a clinician, many of the algorithmically-influenced decisions have already been made. Consent obtained at that point does not cover the decisions that preceded it.

For your most vulnerable patients, children, elderly people, people with cognitive impairments, people from non-English-speaking backgrounds, people who are homeless or in housing crisis, women fleeing family violence, these barriers to meaningful consent are compounded by limited access to information, advocacy, and the institutional channels through which errors are normally identified and corrected.

## 5. What are your governance duties as board directors?

As a board director of a health or social services organisation we carry three categories of duty that are directly engaged by ADM deployment.

### Fiduciary duty

We are required to act in the best interests of those your organisation serves. Approving or continuing the use of an ADM system without understanding its equity implications, how it performs for different patient populations, what happens when it is wrong, how errors are identified and corrected, is not consistent with acting in the best interests of your most vulnerable service users.

### Duty of care

We are required to take reasonable steps to ensure that your organisation's services do not harm those they serve. The research evidence establishes that ADM systems can cause significant harm including, in the healthcare context, increased mortality rates for specific patient groups- that is invisible without explicit monitoring. A duty of care that operates without the information required to detect harm is a duty of care in name only.

### Community trust obligation

Your organisation's relationship with the community depends on the community's confidence that your leadership has patients' long-term interests in mind. This applies during the clinical encounter and also across the full care pathway, not only for the easiest-to-serve patients but for those whose circumstances make them most dependent on the organisation getting it right. ADM deployed without transparency, consent, and accountability mechanisms does not support that confidence. It undermines it.

> *An organisation cannot discharge its governance responsibilities by pointing to the vendor contract when an ADM system causes harm. The community, the regulator, and the law will look to your board. That accountability cannot be contracted out.*

## 6. What should your board do?

There are immediate steps your board should take, and steps to prepare for the regulatory obligations proposed in the submission accompanying this note.

### Steps for board meetings

- Through the Chair ask your CEO, CDO or CRO to identify every system currently in use that contains algorithmic decision-making functionality, including third-party platforms, EMR modules, and decision support tools. Do not accept 'we don't use AI' as a complete answer, ask specifically about decision support features, risk scoring, scheduling optimisation, and clinical trial tools
- Ask whether any of these systems have been assessed for performance across different patient groups and specifically whether performance data exists for the populations your organisation serves who may be underrepresented in training data
- Ask whether any patient currently has a meaningful way to know that an algorithm influenced a decision about their care, or to request human review of such a decision
- Record the board's consideration of ADM governance in the minutes; this establishes that the board has turned its attention to the issue and creates a baseline for future accountability

### Steps to prepare for current and impending regulatory requirements

- Establish an ADM Register; a documented inventory of every system in your organisation containing algorithmic decision-making functionality, with the categories of decision it influences and the patient populations it affects
- Commission an equity performance review of any clinical AI system in use, asking specifically about false negative and false positive rates for socioeconomically disadvantaged, CALD, elderly, and disabled patient groups
- Develop a patient-facing disclosure that explains, in plain language, that your organisation uses AI systems in some aspects of clinical and operational decision-making, and tells patients how to request more information or human review
- Ensure your next platform procurement or upgrade contract explicitly requires the vendor to disclose any algorithmic decision-making functionality, provide subgroup performance data, and notify your organisation of material changes to the system

### Questions to ask at a board meeting

- Which systems in your current clinical and operational environment contain algorithmic decision-making functionality?

- For each of those systems: who approved it, when, and what equity assessment was done before deployment?
- What is the performance of each system for your most vulnerable patient groups compared to the general population?
- What mechanism does a patient currently have to know that an algorithm influenced their care, and to request human review?
- What would happen if one of your AI systems caused demonstrable harm to a patient group? who would know, how quickly, and what would we do?

## 7. The positive is that good governance enables better AI

This note is not arguing against the use of AI in your organisation. Well-governed AI deployment can genuinely improve outcomes for our patients, including and especially for patients whose access to specialist diagnostic capability, personalised care, and timely intervention has historically been limited by cost, geography, and workforce constraints.

Reddy (2026) establishes that AI has real potential to democratise access to healthcare, to make available to a patient in a regional or low-income setting the diagnostic capability that currently requires specialist referral, waiting lists, and travel. In clinical contexts where AI has been rigorously evaluated and introduced with appropriate human oversight; AI-assisted mammography with radiologists retained in the clinical loop, diabetic retinopathy screening with verified referral pathways; the evidence shows genuine improvements in diagnostic accuracy, earlier intervention, and better outcomes.

The governance framework proposed in this submission is the precondition for the beneficial use of AI. An AI system deployed without equity assessment, without patient disclosure, and without outcome monitoring will not reliably deliver its benefits to the patients who need them most because those patients are most likely to be on the wrong side of the system's performance gaps, and least likely to have those gaps identified and corrected.

> *Good governance of AI is not about slowing down innovation. It is about ensuring that the benefits of AI reach the most vulnerable patients rather than compounding the disadvantages they already face. That is both the ethical obligation and the governance responsibility of your board.*

## Summary

If a health service board approves a procurement contract for a new electronic medical record (EMR) or patient flow platform, and that platform contains embedded, unvetted predictive algorithms for patient deterioration or triage, the board is inadvertently bypassing the ACSQHC safety standards and AHPRA guidelines.
My submission to the OAIC is that boards need a synchronized regulatory mandate. Without a clear standard that forces tech vendors to explicitly flag and unbundle these algorithmic components during procurement, public health boards are being exposed to clinical governance and patient safety risks they cannot easily visibility-test.

## Terminology

Automated decision making (ADM): any system using a computer algorithm to generate outputs that influence or determine decisions about individuals regardless of whether it is labelled 'AI', 'machine learning', 'decision support', or any other term.

Algorithmic bias: systematic errors in an AI system's outputs that affect some groups of people differently from others, typically because those groups were underrepresented in the data used to train the system, or because the system uses features that correlate with protected characteristics.

Automation bias: the tendency of clinicians and other professionals to defer to algorithmic outputs without adequate critical scrutiny, particularly when they lack information about the system's limitations.

Double exposure: the compounding dynamic in which the populations most targeted by ADM are simultaneously the populations least equipped to contest algorithmic errors, because the conditions that make them subjects of algorithmic assessment also eliminate their access to correction pathways.

Trust cascade: the chain of information gaps in healthcare ADM governance, from vendor to board to executive to clinician to patient, in which no single ‘actor’ has the visibility required to identify, flag, or correct algorithmic harm.

ADM Register: a documented inventory of every system in an organisation containing algorithmic decision-making functionality, required under the proposed regulatory reforms.

Rights-Based Stakeholder Impact Assessment (RBSIA): a pre-deployment assessment required under the proposed regulatory reforms, evaluating the equity implications, consent architecture, and monitoring mechanisms for any ADM system before it enters clinical or operational use. (or AIA Algorithmic Impact Assessment)