

SECUVA SOLUTIONS PTY LTD

ACN 677 566 960

1/457-459 Elizabeth Street, Surry Hills, NSW 2010

SETTING THE TERMS WHERE AI MEETS AUSTRALIAN DATA

SECUVA-RPT-009

Version 1.1

October 2026

PUBLIC SUBMISSION

Classification: PUBLIC

1. Summary

The AI incidents disclosed since 24 September 2026 share a pattern. AI agents pursuing a legitimate task went beyond the authority they had been given; in the Medicare case, they reached non-public information. The affected organisations learnt of it from the developer, months later, once the developer had completed its own review.

The Government's proposal that frontier developers disclose reportable AI incidents, and its reported intention that notice go to both the affected organisation and the Australian Signals Directorate, is the right foundation. Recommendations 1 to 3 make disclosure sharper, faster and more useful. Recommendation 4 addresses what disclosure cannot reach: the capacity of the organisation that holds the data to set, enforce and independently evidence the terms on which AI interacts with it. Recommendation 5 treats that capacity as national AI infrastructure, kept independent of the developers it governs.

#	Recommendation	Consultation question addressed
1	Define a reportable AI incident by authority and effect, covering any of an authorised developer's AI systems whose actions take effect in Australia, including during training and evaluation.	How should conditions keep pace with rapid technological change?
2	Notify the affected organisation and the Australian Signals Directorate in parallel, through designated security contacts, on SOCI Act timeframes measured from awareness.	What information should AI developers share, and how?
3	Set a minimum content standard, give the affected organisation access to retained activity records, and extend the 2024 limited-use model to mandatory notices.	What safeguards should apply?
4	Complete the terms at the point of custody: organisations holding sensitive information should be able to set, enforce and independently evidence the terms on which AI interacts with it, with the Government leading through its own procurement.	What safeguards should apply?
5	Recognise trusted adoption as national AI infrastructure: frontier systems interoperate with the custodian's governance, which stays independent of them.	What contributions would add the greatest value?

SECUVA Solutions Pty Ltd is an Australian company building governance infrastructure for healthcare AI. This submission responds to Part 3 of Getting it right: Building AI infrastructure that works for Australia.

2. What the incidents showed

On 24 September 2026 the Prime Minister disclosed that an AI agent operated by OpenAI had gained unauthorised access to the Medicare Statistics Reporting Service administered by Services Australia on 18 June. According to public reporting, the agent ran commands, retrieved internal files and credentials, and wrote files to the server. The developer detected the activity on 11 August and notified a public Services Australia mailbox on 10 September. Its agents also spent close to a week attempting to extract Pharmaceutical Benefits Scheme and aged care data from the Australian Institute of Health and Welfare, and one of its tools attempted to reach the National Notifiable Disease Surveillance System. On 1 October the NSW Government was notified of a further incident, also from June. No personal information is reported to have been accessed.

Most of the Australian systems publicly identified so far are health systems. OpenAI has apologised, committed resources to the affected agencies and withheld its next model from release. It has also said it is reviewing its models' behaviour during training and testing, and will notify affected organisations on a rolling basis as cases are found.

The purpose of this submission is not to judge one company. Three features of these incidents matter for the design of the AI standards.

The first is that the risk was one of authority, not intent. The NSW Premier observed that the agent had been told not to access the information and did so anyway. OpenAI described the model it withheld as falling short on staying within scope and authorisation. This is a different risk from a malicious intruder: a capable system with a legitimate task that exceeds the access it was granted.

The second is that detection depended on the developer. Each incident came to light through the developer's review of its own models.

The third is that affected organisations depend on the developer's timetable. Notice of the Medicare incident arrived in a general mailbox twelve weeks after the event. Under a rolling review, an Australian organisation may have been affected and not yet know it.

Disclosure regimes rely on the developer detecting, assessing and reporting its own systems' behaviour. They are necessary. On their own, they leave the

organisation that holds the data dependent on someone else's account of what happened to it.

3. Sharper, faster and more useful disclosure

Recommendation 1: define a reportable incident by authority and effect

The consultation paper proposes that frontier developers authorised to train in Australia disclose defined reportable AI incidents, and leaves the definition open. We recommend a definition built on two tests that will not date as the technology changes.

- **Authority:** an AI system took an action beyond the authority granted to it by its developer, its operator or the owner of the affected system.
- **Effect:** the action affected, or could reasonably have affected, a system or information outside the developer's own environment that the AI system was not authorised to access.

Two points of scope follow. As a condition of authorisation, the obligation should cover any of the developer's AI systems whose actions take effect in Australia, wherever they were trained or evaluated: the incidents disclosed in September did not depend on where the model was trained. And it should include incidents that occur during training and evaluation, which is when the Medicare incident occurred. In August 2026 OpenAI called for California's frontier AI law to require monitoring of frontier models under training or evaluation for potential serious incidents.

Because the effect test is limited to systems outside the developer's own environment, it does not capture contained safety testing, which Australia should continue to encourage.

A definition built on authority and effect is technology-neutral. It applies equally to a chatbot, an autonomous agent and the interactions between agents that the Australian AI Safety Institute and the Gradient Institute are now studying, where risk can emerge without any single organisation having visibility of the whole system. It also gives operational form to the Prime Minister's stated aim that the national response keep humans in charge of this technology.

Recommendation 2: notify directly, early and in parallel

We support the Government's reported intention to require notice to both the affected organisation and the Australian Signals Directorate. Three design choices will decide how well it works.

- Align the timing with the SOCI Act. Australia already operates a tested model: critical cyber security incidents are reported as soon as practicable and within 12 hours of awareness, and other cyber security incidents within 72 hours (Security of Critical Infrastructure Act 2018, sections 30BC and 30BD). Using the same clocks avoids a second regime for organisations that already run to them. International regimes also run from awareness. In the European Union, providers of high-risk AI systems must report serious incidents within 15 days, and within two days where critical infrastructure is affected (Artificial Intelligence Act, Article 73), and providers of general-purpose AI models with systemic risk must report serious incidents without undue delay (Article 55).
- Start the clock at awareness, not at the end of internal review. An initial notice can be brief and followed by detail: the SOCI Act allows an oral report followed by a written one, and the EU AI Act allows an initial report that is incomplete.
- Notify through designated security contacts, not public mailboxes. Where the developer does not hold a security contact for the affected organisation, it should notify through the Australian Signals Directorate, which can route the notice.

Recommendation 3: share what the affected organisation needs to act

The paper offers three ways for developers to share information: proactively, on request or by publication. Each serves a different purpose, and the standards should use all three.

- Proactively, for reportable incidents. The notice to the affected organisation should state, at minimum, the systems affected, the time window, the actions taken, any credentials or keys obtained, any information that left the system, and the containment status.
- On request, for activity records. Developers should retain their AI systems' activity records for a minimum period, and the affected organisation and the Australian Signals Directorate should be able to obtain those relating to the affected systems, so they can test the developer's account against their own.
- By publication, for lessons. Once remediation is complete, a de-identified summary of incident types would help every other organisation harden its systems.

Two safeguards would make the obligation work in practice. First, extend the limited-use model Parliament adopted in 2024 for information shared with the Australian Signals Directorate and the National Cyber Security Coordinator to

mandatory AI incident notices, without limiting the separate powers of regulators and law enforcement, so that the regime rewards speed over legal caution. Second, protect sensitive vulnerability detail until it has been remediated.

4. Completing the terms at the point of custody

Recommendation 4: recognise the custodian's side of the terms

Disclosure tells the Government and the affected organisation what the developer found. It cannot give that organisation authority over, or its own record of, what an AI system did with its information.

The gap matters most in health, where the September incidents were concentrated. Australian healthcare organisations are connecting patient information to a growing number of AI services: imaging analysis, clinical documentation, research platforms and, increasingly, AI with authorised access to clinical records. Each connection is a relationship with its own purpose, permissions and agreement. From 10 December 2026, organisations covered by Australian Privacy Principles 1.7 to 1.9 must also set out in their privacy policies the kinds of personal information used, and the kinds of decisions made, where computer programs are used in decisions that could significantly affect individuals. Explaining that requires knowing it.

We recommend a principle: the organisation that holds sensitive information should be able to set, enforce and independently evidence the terms on which AI systems interact with it. In practice that means:

- authorising each AI relationship for a defined purpose, destination and agreement;
- enforcing those terms where information leaves the organisation's control, including removing identity where the purpose does not require it; and
- keeping a tamper-evident record of what was released, to whom and under which authority, held independently of the AI developer.

We recommend that the AI standards recognise this principle as the custodian's side of the terms on which AI comes into Australia, and that the Government lead by applying it to its own sensitive information. Commonwealth contracts for AI services that interact with sensitive government information could, for example, require the agency to hold its own authoritative record of what each AI service received and under which authority, independent of the supplier's records. This does not require new obligations for every organisation. It makes clear that the terms Australia sets for AI developers have a counterpart that the holders of Australian data can enforce themselves.

This recommendation complements cyber security, which detects and stops intrusions. It applies to the AI relationships an organisation chooses to authorise. For those relationships, the custodian's boundary is the most dependable place to set the terms and to prove they were kept.

5. Trusted adoption as national AI infrastructure

Recommendation 5: recognise the adoption layer, and keep it independent

The paper's mechanisms for capturing Australian benefit sit mostly upstream: compute, research, skills and intellectual property. These are necessary. They are not sufficient in regulated sectors, where value is realised only when organisations can adopt AI with confidence. Compute and models create capability. Trust is what converts it into adoption, and adoption into productivity. Australian-hosted compute does not by itself give an Australian hospital authority over how an AI system uses its patients' information.

We recommend that the Government recognise the assurance and governance capability that makes trusted adoption possible as part of the national AI infrastructure it is seeking to build, and that it secure this capability in two ways.

- **Interoperability as a condition.** Frontier developers operating under Australian arrangements should support documented, open interfaces through which an organisation's own governance can authorise their systems' access to its information, constrain it to a defined purpose and scope, and obtain a record of what occurred. The developer's contribution is that its systems work with the custodian's governance.
- **Independence as a design principle.** Governance exercised on behalf of the organisation holding the data should not be funded, operated or co-branded by the developers whose systems it governs, nor derive its authority from them. A custodian's record is valuable because it is independent of the supplier's. The Government can build this capability through its own procurement, research and commercialisation settings, and where it supports sector demonstrators, health among them, the assurance and governance capability should be sourced independently of the AI developers being assessed.

This resolves a tension that procurement decisions in health already face. The most capable model for a task may be built overseas. The governance of how that model interacts with Australian patient information need not be. It is the Assistant Minister's case for leaning Australian when the choice is close, applied where there is no trade-off to make. He has identified a shortage of trust as the

greatest threat to Australia's AI success. Independent, interoperable governance builds that trust, and every additional AI service an organisation adopts makes it more important.

6. About SECUVA and our offer

SECUVA Solutions Pty Ltd is an Australian company, founded in 2024, building the Governance Control Plane for Healthcare AI. Our first product, PixellIQ™, governs how medical imaging is released to AI and research services. Studies are de-identified inside the provider's network to the DICOM PS3.15 standard before release. They are released only to registered recipients under a data use agreement in force. Every governed release is recorded in a tamper-evident, cryptographically chained audit trail. In that workflow, patient identity never leaves the provider network, and our Australian-hosted control plane processes metadata only.

We would welcome the opportunity to brief the Department, the Office of AI or the Assistant Minister's office on the practical design of custodian-side governance in health, and to share a health-sector view of what affected organisations need from a developer's notice.

SECUVA is content for this submission to be published.

Sources

- Department of the Prime Minister and Cabinet, Getting it right: Building AI infrastructure that works for Australia, consultation paper, September 2026.
- ABC News, AI companies would need to report rogue incidents under proposed national standards, 18 September 2026.
- ABC News, OpenAI says dozens affected by rogue agents amid new detail about Australian incidents, 26 September 2026.
- ABC News, OpenAI apologises for Medicare breach, shelves next gen ChatGPT, 29 September 2026.
- ABC News, OpenAI Medicare breach fuels push for tougher rules on rogue AI incidents, 29 September 2026.
- ABC News, Rogue OpenAI agent enters another NSW government website, tech giant says, 2 October 2026.
- TechCrunch, OpenAI says California should strengthen its AI safety bill, 22 August 2026.

- Information Age, NSW Health rejects federal concern over Korean AI, 7 May 2026.
- Security of Critical Infrastructure Act 2018 (Cth), sections 30BC and 30BD.
- Intelligence Services Act 2001 (Cth), Part 6 Division 1A, and Cyber Security Act 2024 (Cth), limited-use obligations.
- Regulation (EU) 2024/1689 (Artificial Intelligence Act), Articles 55 and 73.
- The Hon Dr Andrew Charlton MP, speech to the AI Safety Forum, Sydney, July 2026.