Setting the Standard: An Australian Model for AI Incident Monitoring & Reporting

October 2026

Introduction and Executive Summary

On 18 June, an OpenAI agent autonomously gained access to non-public information through a Services Australia portal. OpenAI became aware of this incident in August, but did not notify the Australian Government until 10 September (and used a generic public email to provide that notification).

As AI becomes more capable and autonomous, incidents like this will become more common. Yet Australia has little visibility over when they occur: we only find out about breaches when AI companies choose to disclose them.

To fix this gap, Australia needs an AI incident monitoring and reporting (AIIMR) regime.

Global Shield Australia has been calling for a legislated AIIMR regime in Australia since the 2025 election. Such a system would track and enable responses to significant AI failures, harms, security incidents, and near misses, just as we have in aviation and other industries.

There are several models overseas and in other sectors that can be drawn from when developing Australia's AIIMR regime. However, these cannot just be duplicated and applied to AI in the Australian context. Recent AI incidents have demonstrated gaps in AIIMR systems overseas, and shown where they can be built upon to meet the expectations of Australians.[1]

Global Shield Australia has prepared a detailed proposal setting out the key elements needed for an effective and world-leading Australian AIIMR regime (see **Appendix A**). This work is informed by practice overseas, existing non-AI incident reporting schemes within Australia, and expert analysis and commentary.

The key elements of the proposed approach are set out in **Table 1**.

As Australia looks to establish an AIIMR regime, getting its design right will be critical to ensuring it provides the government with real intelligence on AI risk, enables responses to threats and harms, and empowers Australians harmed by AI. Australia should seize this opportunity to lead and set a new standard for our region and the world more broadly to follow.

Contact: australia@globalshieldpolicy.org

[1] See, e.g., Z. Kara, *Improving Frontier AI Incident Reporting Regimes* (GovAI, Policy Brief, September 2026).

Table 1. Six key elements of an effective AIIMR regime
1. Purpose
The AIIMR regime must have a clear purpose to ensure a coherent approach is taken to matters such as what incidents are captured and what powers the receiving agency has to respond. The aims of Australia's AIIMR regime should be to: (a) Activate incident and emergency response mechanisms to minimise harms and contain incidents; (b) Identify factors that contribute to incidents; (c) Ensure appropriate corrective and remediation action is undertaken; and (d) Enable learning and improved risk controls.
2. Scope
The AIIMR regime should apply to entities that have visibility of AI incidents from a range of perspectives. Not all AI models pose the same level of risk. The AIIMR regime should focus on the most capable, highest-risk AI models and deployments. Australia's AIIMR obligations should apply to at least three categories of entities: (1) companies training and serving frontier AI models (identified by compute used, breadth of training, assessed capabilities and risk, and Ministerial designation), (2) data centres hosting the training or inference of such models, and (3) third parties hosting or testing such models. The obligations should apply during training, internal use, and deployment. Voluntary reporting pathways should also be available for users of AI systems, those impacted by incidents, and whistleblowers.
3. Monitoring and logging obligations
To ensure that reporting obligations do not incentivise entities to avoid monitoring for incidents, it is important to require them to actively monitor and log their systems for reportable incidents. Australia's AIIMR regime should require covered entities to have comprehensive and effective monitoring and logging systems in place. The regime should specify what records need to be preserved to enable proper analysis of incidents when they happen, and ensure that actions can be attributed to relevant AI agents.
4. Reportable incidents
The AIIMR regime should capture clearly defined harms. It should also apply to incidents that demonstrate a heightened risk. The regime should not adopt unrealistically high thresholds set in overseas regimes, but be calibrated against existing regimes in other Australian industries. Australia's AIIMR regime should require covered entities to report incidents defined across three categories: (a) **Serious harms:** where a frontier AI model, as a result of misuse, malfunction or autonomous action, causes, materially contributes to, or is reasonably likely to have

caused or contributed to: one or more deaths or serious injuries, serious criminal offences, a serious cyber security incident, critical infrastructure disruption, serious property or economic damage, serious harm to the environment, or prejudice to Australia's national security;

(b) **Model, oversight, and security failures (regardless of whether harm results):** loss of control, subversion of monitoring (or credible attempts at subversion), acting contrary to instructions, unauthorised access to or exfiltration of model weights (including by the model itself), insider threats, and other serious cyber security incidents targeted at a model or its host; and

(c) **Near misses:** where any of the above incidents are avoided by an unplanned intervention or chance.

Further guidance on thresholds and definitions will be needed for many of the categories above to ensure clarity for industry.

5. Timeframes for reporting

Consistent with similar regimes overseas and in other sectors, reporting deadlines should be tiered based on severity or likelihood of harm.

An initial report should be required within: **twenty-four hours**, where there is an imminent risk of, or actual, death or serious injury; **forty-eight hours,** where there is a widespread or systemic risk, including of critical infrastructure disruption; and **seventy-two hours,** for all other reportable incidents.

Comprehensive reports could then also follow **within 15 days**, providing entities sufficient time to collect and analyse all relevant information. Final reports, lodged **within 30 days** of the incident being resolved, should include causal analyses and remediation measures undertaken.

6. The receiving agency

The agency receiving AI incident reports must have sufficient technical capabilities to understand and act upon reports; broad enough expertise to address the variety of potential threat models (not just cyber incidents); the ability to actively engage with a range of stakeholders.

To be effective, the AIIMR agency should have the powers to:

(a) Compel information, documents, testimony, and access to models and relevant model records (such as chain-of-thought transcripts or evaluation logs);

(b) Conduct or order independent investigations of sufficiently serious incidents (including near misses);

(c) Issue safety bulletins and recommendations to industry and users; and

(d) Require corrective action plans or direct covered entities to restrict, modify or withdraw a model, or take other action to protect the public or respond to an incident.

The AIIMR agency should also share incident reports with relevant regulators for response.

Box 1. How an AIIMR would have applied to the Services Australia OpenAI incident

On 18 June 2026, an unreleased OpenAI agent undergoing testing circumvented blocks on Services Australia's Medicare Statistics Reporting Service website and accessed non-public information. OpenAI discovered the incident in August and reported it to the Australian Government on 10 September via a public Services Australia email address. ASD was informed of the incident on 15 September, and the broader public on 24 September.[2]

If Australia had an AIIMR of the kind described in this submission:

(a) **OpenAI would likely have identified the incident sooner:** OpenAI would have been required to effectively log and monitor the actions of its agents (including during testing), making it more likely that the incident would have been identified in June, not months later in August.

(b) **OpenAI would have been legally obliged to report the incident:** the incident would likely have been reportable on two bases. First, as the agent acted beyond its tasking and breached OpenAI's guardrails. Second, as it appears to have gained unauthorised access to an Australian government system.

(c) **The government would have been notified within 72 hours:** OpenAI would have been required to provide an initial report of the incident within 72 hours. This would have gone through a formal reporting channel to the AIIMR agency, which could then refer the matter to ASD and Services Australia.

(d) The government would have had powers to investigate and issue warnings to others: the AIIMR agency could have requested preserved logs relating to the incident, ordered a more detailed investigation, and notified other agencies (or had ASD notify other agencies) of the need to address similar issues in their publicly facing services.

(e) **OpenAI would have been obliged to provide a full explanation for what went wrong and what corrective action had been taken:** comprehensive and final reports would have followed OpenAI's initial report. These would set out the company's analysis of the causes of the incident, how the company responded, and what actions were taken to prevent recurrence of the same issue in the future.

(f) **There would be broader learning:** de-identified lessons from the incident could be published to improve the safety and security of the broader cyber and AI ecosystem.

This would have improved the response in this specific case. It would also be all the more important where the harm is far more severe, such as if an AI system is being used to plan a terrorist attack or for biological weapons development. In such cases, it is crucial that the government is promptly informed, has the powers to investigate and respond, and then has mechanisms in place to ensure lessons learned are quickly promulgated.

[2] ABC News, *OpenAI hacked Medicare portal, Prime Minister Anthony Albanese says* (24 Sep 2026); OpenAI, *How We Will Do Better for Australia* (28 Sep 2026).

Appendix A: An Australian AI Incident Monitoring and Reporting Regime

The Consultation Paper recognises that AI incident reporting is an important part of the regulatory toolkit that Australia needs to ensure the development of a safe and secure Australian AI sector. Getting the design of such a reporting regime right will be key, including ensuring all relevant incidents are captured and reported, that the government has appropriate powers to respond to and learn from incidents, and to build upon lessons learned from approaches taken overseas.

This Appendix sets out the key elements of an AI Incident Monitoring and Reporting (**AIIMR**) regime with recommendations for the approach Australia should take.

	Element & description	An Australian approach
1. Purpose		
1.1	**Purpose of the regime** *The AIIMR regime must have a clear purpose to ensure a coherent approach is taken to its various elements, such as what incidents are captured, how reports are dealt with, and what powers the receiving agency has to respond.*	An Australian AIIMR should have four purposes: (a) Activating incident and emergency response mechanisms to minimise harms and contain incidents; (b) Identifying factors that contribute to incidents; (c) Ensuring appropriate corrective and remediation action is undertaken; and (d) Enabling learning and improved risk controls.
	Precedents: *Cyber Security Act 2024, s 3 (aims include to facilitate response to incidents, prevent and improve the detection of incidents, and facilitate sharing of information about incidents); Transport Safety Investigation Act 2003, s 12AA (functions include to identify factors that contribute to safety matters and publicly report on investigations); AI Act (EU), recital 155 (aims include to ensure AI providers can learn from experience to improve systems, designs, and development, and take timely response action).*	
2. Scope — who and what is covered		
2.1	**Covered AI models** *Not all AI models pose the same level of risk. Limiting the AIIMR regime's coverage will minimise regulatory burden and ensure the system is not overwhelmed.*	Frontier AI models with coverage determined by: (a) Compute used in their training (e.g. $>10^{26}$ FLOPs); (b) The breadth of their training data; (c) Their capabilities, generality, adaptability, and assessed level of risk; and/or (d) Ministerial designation. These thresholds should be set in regulations, not statute, to enable them to be regularly reviewed and updated.

	Element & description	An Australian approach
		The Consultation Paper contemplates applying incident reporting to frontier AI training. However, to be effective, this obligation should continue to apply to both internal use and external deployments. It must apply throughout the AI model's life cycle, especially when the product, service, or model is actively available in the Australian market.
	Precedents: *Transparency in Frontier Artificial Intelligence Act (California), §22757.11(f), (i), (j) (referring to compute, breadth of training data, generality of output, adaptability, and revenue); AI Act (EU), arts 3(63)-(65), 51 (referring to training data size, capabilities, and compute); AI Basic Act (Korea), s 32(1) (allowing compute thresholds to be set by Presidential Decree); Security of Critical Infrastructure Act, s 51 (enabling Ministers to declare assets to be critical infrastructure assets).*	
2.2	**Entities required to monitor and report** *Mandatory monitoring and reporting obligations should apply to those entities with access to and visibility of AI incidents and their impacts.*	Mandatory monitoring and reporting obligations should apply to: (a) Companies training or deploying frontier AI models in Australia or serving such models to Australian users; (b) Australian data centres hosting frontier AI training or inference, in regard to incidents they can observe (e.g. security incidents; unauthorised access/use); and (c) Third parties evaluating, training, or hosting frontier AI models. Pathways for voluntary reports should also be available from a broad range of stakeholders and users.
	Precedents: *Transparency in Frontier Artificial Intelligence Act (California), §22757.11(h), (j) (defining frontier developer); Responsible AI Safety and Education Act (New York), §1425 (specifying the Act applies to frontier models developed, deployed, or operating in New York); AI Act (EU), art 3(3), (4) (defining provider and deployer); Transport Safety Investigation (Voluntary and Confidential Reporting Scheme) Regulation 2012 (REPCON) (voluntary transport safety incident reporting regime); Security of Critical Infrastructure Act, Part 2B (cyber security incident notification requirements).* *See also C. Bullock et al., Legal Considerations for Defining "Frontier Model" (LawAI, Working Paper No. 2-2024, 2024); and M. Pistillo et al., The Role of Compute Thresholds for AI Governance, 1 George Washington Journal of Law and Technology 26 (2025).*	
2.3	**Reportable incidents** *The AIIMR regime should capture clearly defined harms where these have actually occurred or have been narrowly avoided. It should also apply to specific occurrences and events that indicate heightened risk.*	Reportable incidents should be clearly defined and objectively determinable events. A reportable incident should be defined as the occurrence of an event in which an AI model, agent, product, or service, whether through misuse or malfunction (including misaligned function), caused, materially contributed to, or is

	Element & description	An Australian approach
		reasonably likely to have caused or contributed to: (a) One or more deaths or serious injuries; (b) The commission of serious criminal offences, including development or deployment of chemical, biological, radiological, or nuclear weapons, or cyber attacks (whether by a human or AI); (c) A serious cyber security incident, particularly involving government systems; (d) Disruption to critical infrastructure; (e) Serious damage to property or economic interests (with defined thresholds); (f) Serious harm to the environment (with defined thresholds); or (g) Prejudice to Australia's national security (with clear guidance to industry). Reportable incidents should also include the following, whether or not harm results: (a) Loss of control of a frontier AI model; (b) A frontier AI model subverting, or credibly attempting to subvert, monitoring or controls, or acting materially outside of or contrary to its instructions (except when being evaluated to specifically test such behaviours, with stringent controls applied); (c) Unauthorised access to, or exfiltration of, frontier AI model weights, or a credible attempt to do so (whether by a human or AI); (d) Insider threats such as employee espionage or misuse of or tampering with models and internal data; and (e) Serious security and cyber incidents targeting a frontier AI model or host data centre. Reportable incidents should also include near misses, where any of the above harms or occurrences was avoided by an unplanned intervention or through chance.
	Precedents: *Transparency in Frontier Artificial Intelligence Act (California), §22757.11(c), (d) (definitions of*	

	Element & description	An Australian approach
	'catastrophic risk' and 'critical safety incident'); AI Act (EU), art 3 (definition of 'serious incident'); Law on Artificial Intelligence (Vietnam), art 3(8) (definition of 'serious incident'); Proposed AI Incident Reporting Act HR 9477 (US), s 2(b)(2) (definition of 'reportable activity'); Transport Safety Investigation Regulations 2021, regs 6 and 11A-12 (definitions of 'aircraft accident', 'aircraft incident' and 'reportable matters'); Work Health and Safety Act 2011, s 37 (defining 'dangerous incident'); National Health Security Act 2007, s 48 (definition of 'reportable events'); Therapeutic Goods Act 1989, ss 29A and 29AA (reportable information); Therapeutic Goods (Medical Devices) Regulations 2002, reg 5.7 (coverage of near misses); Australian Consumer Law, s 131 (definition of reportable incidents).	
2.4	**Exemptions** *Exemptions may be necessary to exclude legitimate use cases where reporting would be counterproductive, such as in academic research, defence and national security, with clearly defined limits.*	Any exemptions to Australia's AIIMR regime should be minimised and narrowly scoped. Two categories for potential exemptions are: (a) Non-commercial research by universities and public research bodies for the purposes of understanding existing deployed models, except where engaged by a frontier AI company as a third-party evaluator; (b) Models explicitly contracted for development and use by Australian government entities for national security and defence purposes. For national security and defence exemptions, there should be separate monitoring and reporting obligations at least as stringent as those that would apply under the AIIMR regime.
	Precedents: *Responsible AI Safety and Education Act (New York), §1426 (academic research exemption); AI Act (EU), art 2(3), (6) (academic, national security and defence exemptions); AI Basic Act (Korea), art 4(2) (national security and defence exemption).*	
3. Monitoring and reporting — how and when		
3.1	**Duty to monitor and log** *To ensure that reporting obligations do not incentivise entities to avoid monitoring for incidents, it is important to require covered entities to actively monitor and log their systems for potential reportable incidents.*	Covered entities should be required to document and maintain a system to monitor, detect, and record reportable incidents. The system should be proportionate to the entity's size and role, the relevant frontier AI model's capabilities, and how it is being used. This system should also assess suspected reportable incidents and record whether or not they were judged as reportable and why. The system should cover training, internal use, and deployment. It should log actions of AI agents, chain of thought and other reasoning transcripts, and reports from users, staff, and external researchers. This logging should include the ability to attribute actions to specific models and agent instances. It should actively monitor for harms and

	Element & description	An Australian approach
		potential incidents during deployment, including through interactions with other AI and computer systems. Logs and records created by AI agents also need to be protected from AI agents, given past attempts by AI agents to tamper with these records.
	Precedents: *AI Act (EU), arts 55(1)(c), 72 (obligation to monitor and document); Code of Practice for General-Purpose AI Models (Safety and Security Chapter) (EU), Measure 9.2 (relevant information for tracking serious incidents); Transparency in Frontier Artificial Intelligence Act (California), §22757.12(a)(8) (obligation to publish framework including for identifying critical safety incidents); AI Basic Act (Korea), art 32(1) (obligation to have a system that monitors safety accidents); APRA CPS 234, para. 23 (obligation to have a security incident detection mechanism); APRA CPS 230, paras. 16, 32-33 (obligation to identify operational risk incidents and near misses).* *See also Z. Kara, Improving Frontier AI Incident Reporting Regimes (GovAI, Policy Brief, September 2026), pp. 10-12.*	
3.2	**Agency receiving reports** *The agency receiving AI incident reports must have sufficient technical capabilities to understand and act upon reports. It must also have broad enough expertise to address the variety of potential threat models or refer reports to relevant regulators.*	Australia's AI Safety Institute (**AISI**) has the appropriate technical and governance capabilities to assess, refer, and act on AI incident reports. It is also connected to the international work of AISIs, enabling information and threat sharing. However, incorporating the AISI into an AIIMR will require providing it with greater statutory backing and resourcing. The Australian Signals Directorate's (**ASD**) Australian Cyber Security Centre (**ACSC**) is also a potential option for receiving reports, noting its existing role receiving cyber reports.[3] However, its integration within Australia's intelligence community may limit its ability or willingness to share reports more broadly. Its particular expertise in cyber security may also make it a less appropriate recipient of reports of broader AI model incidents (such as those relating to biological risk or harms to human health). Alternatively, a new standalone body could be made responsible for administering the AIIMR regime.
	Precedents: *AI Act (EU), arts 55 and 73 (AI Office and market surveillance authorities); Transparency in Frontier Artificial Intelligence Act (California), §22757.13 (Office of Emergency Services); Responsible AI Safety and Education Act (New York), §1420(16) (an office within the Department of Financial Services).*	

[3] Australian Signals Directorate, *Report a cyber security incident*.

	Element & description	An Australian approach
3.3	**Timing of reports** *Consistent with similar regimes overseas and in other sectors, reporting deadlines should be tiered based on severity or likelihood of harm.*	Reporting timeframes should commence when a covered entity has facts sufficient to support a reasonable belief that a reportable incident has occurred. Specific timeframes may need to be co-designed with industry. However, an initial report could be required within: (a) **Twenty-four hours**, where there is an imminent risk of, or actual, death or serious injury; (b) **Forty-eight hours,** where there is a widespread or systemic risk, including of critical infrastructure disruption; (c) **Seventy-two hours,** for all other reportable incidents. Incidents could also be made reportable immediately or as soon as reasonably practicable where harm is imminent or has occurred. Comprehensive reports could then also be required to follow **within 15 days**, providing entities sufficient time to collect and analyse all relevant information. Final reports, lodged **within 30 days** of the incident being resolved, should include causal analyses and remediation measures undertaken.
	Precedents: *AI Act (EU), art 73 (initial report within 15 days, 2 days for widespread infringement, 10 days for death); Transparency in Frontier Artificial Intelligence Act (California), §22757.13(c) (initial report within 15 days or 24 hours to authorities if imminent risk of death or serious injury); Responsible AI Safety and Education Act (New York), §1422(3) (initial report in 72 hours or 24 hours to authorities if imminent risk of death or serious injury); Security of Critical Infrastructure Act 2018, ss 30BC-30BD (cyber incident reports within 12 or 72 hours depending on severity).*	
3.4	**Contents of reports** *Standard form reporting will help ensure reports are comparable, able to be easily analysed and shared, and that industry is clear on what is required. These should also be interoperable with overseas regimes (to the extent possible).*	The precise contents of reports should be co-designed with industry. The contents will also vary between initial, comprehensive and final reports. Common categories of reportable information for incidents, including under OECD guidance, have included: (a) **Metadata:** title, description, date, countries concerned, submitter details; (b) **Technical information:** model name and

	Element & description	An Australian approach
		version, agent identifiers, developer information, extent of deployment; (c) **Harm:** type, scale, involvement of the AI model, and why the incident is reportable; (d) **Affected people, entities or systems**; and (e) **Remediation undertaken or planned**. Access to system logs, transcripts, or the model itself may also be required in some circumstances. These should be preserved and access enabled upon request. The occurrence of an incident should also trigger preservation obligations to suspend ordinary deletion or overwriting procedures of relevant records.
	Precedents: *Incident Report for Serious Incidents under the AI Act (High-risk AI systems) (Draft Reporting Template) (EU); Report for Serious Incidents under the AI Act (General-Purpose AI Models with Systemic Risk) (Draft Reporting Template) (EU).* *See also Z. Kara, Improving Frontier AI Incident Reporting Regimes (GovAI, Policy Brief, September 2026), 13; and OECD, Towards a common reporting framework for AI incidents (2025).*	
3.5	**Regulatory overlap** *An AIIMR will overlap with other regulatory regimes, such as for critical infrastructure operators, or in relation to cyber security incidents. Duplicate reporting should be minimised to avoid unnecessary costs and inconsistent responses.*	Where reporting entities are covered by equivalent reporting requirements, compliance with those requirements should be deemed compliance with the AIIMR. To provide clarity and certainty, the equivalent regimes should be specified in rules or a Ministerial determination. This should be supported by information-sharing between regulators to ensure there is a central source compiling and tracking AI incidents in particular.
	Precedents: *AI Act (EU), arts 73(9)-(10); Transparency in Frontier Artificial Intelligence Act (California), §22757.13(h)-(i); Responsible AI Safety and Education Act (New York), §1422(9); Australian Consumer Law, s 131(2)(c).*	
4. Response — how reports are used		
4.1	**Use of reports** *There should be clear guidance on what incident reports will be used for, in line with the purpose of the AIIMR regime.*	AI incident reports should be primarily used for three purposes: (a) Referral to response agencies to activate emergency and regulatory responses; (b) Review of significant incidents to recommend mitigations to avoid recurrence or escalation of harms in future, including through investigation or ordered investigation of incidents; and

	Element & description	An Australian approach
		(c) Remediation of incidents, including through the use of specific powers to direct remediation and prevention action by the reporting entity. Reporting should not grant safe harbour, and other regulators should be able use the content of shared reports to implement other statutes.
	Precedents: *AI Act (EU), arts 73(7)-(8) (referral of reports to authorities and intervention where necessary); Transparency in Frontier Artificial Intelligence Act (California), §22757.13(e), (g) (referral of reports to agencies and public reporting); Transport Safety Investigation Act 2003, ss 25, 25A (publication of reports on incidents and recommendation of safety actions).*	
4.2	**Response powers** *To enable effective remediation and response, new powers will be needed for the agency responsible for handling AI incident reports.*	The agency receiving AI incident reports will need powers to: (a) Compel information, documents, testimony, and access to models and relevant model records (such as chain-of-thought transcripts or evaluation logs); (b) Conduct or order independent investigations of sufficiently serious incidents (including near misses); (c) Issue safety bulletins and recommendations to industry and users; and (d) Require corrective action plans or direct covered entities to restrict, modify or withdraw a model, or take other action to protect the public or respond to an incident.
	Precedents: *AI Act (EU), arts 91-93 (powers to request information, evaluate AI models, request AI companies take certain measures); AI Basic Act (Korea), art 40 (power to order data production, conduct investigations, or order corrective actions); Law on Artificial Intelligence (Vietnam), art 12(3) (power to order suspension or recall); Transport Safety Investigation Act 2003, Part 5 (investigation powers); Security of Critical Infrastructure Act 2018, Part 3A (intervention powers); Australian Consumer Law, ss 122, 129 (recall and safety warning notice powers).*	
4.3	**Information sharing** *AI incidents will often cut across sectors and regulators. Information sharing between regulators, and with partners overseas, will be needed to ensure an accurate picture of the threat landscape can be constructed.*	The Australian AIIMR agency should be able to: (a) Share relevant reports with other agencies and regulators, with appropriate protections for commercial-in-confidence data or information; (b) Share reports and analysis with overseas regulators and AISIs, subject to confidentiality arrangements where necessary; (c) Produce public reporting of de-identified

	Element & description	An Australian approach
		incidents and lessons learned; and (d) Provide annual reports to Parliament on incidents, responses, and trends.
	Precedents: *AI Act (EU), art 78(5) (ability to exchange information with other countries under confidentiality arrangements); Transparency in Frontier Artificial Intelligence Act (California), §22757.13(g) (annual reports); Therapeutic Goods Administration's Database of Adverse Event Notifications; Office of the Australian Information Commissioner's Notifiable Data Breaches Publications.*	
5. Protections and enforcement		
5.1	**Confidentiality and limits on use of reports** *Reporting entities should be incentivised to report fully, honestly and swiftly. This requires protections for confidential information and against self-incrimination, while not preventing legitimate regulatory action against entities.*	Australia's AIIMR regime should consider drawing on precedents in similar regimes, such as: (a) Exemptions from freedom of information and other disclosure requirements; (b) Protections that maintain legal professional privilege over disclosed materials; and (c) Clarification that AI incident reports are not an admission of fault or liability.
	Precedents: *AI Act (EU), art 78(5) (confidentiality protections); Transparency in Frontier Artificial Intelligence Act (California), §22757.13(f) (public records exemption); Cyber Security Act 2024, s 31 (protections for legal professional privilege).*	
5.2	**Whistleblower protections** *Protections for whistleblowers can help incentivise reporting entities to develop a culture of reporting by preventing retaliation against staff reporting failures or harms.*	Australia has existing whistleblower protections, particularly in relation to misconduct. These should be reviewed to determine if amendments are needed to ensure they provide sufficient protections for reports regarding AI incidents, such as: (a) Protection from retaliation from the reporting entity; (b) Requiring reporting entities to have internal reporting channels; (c) Voiding contractual terms that prevent disclosures; and (d) Potential protection from prosecution for certain categories of reports.
	Precedents: *AI Act (EU), art 87 (application of EU whistleblowing directive to the AI Act); Labor Code (California), §1107.1 (whistleblower protections for frontier AI company employees); Corporations Act 2001, Part 9.4AAA (corporate whistleblower protections); Public Interest Disclosure Act 2013 (public service whistleblower protections).*	
5.3	**Penalties and enforcement** *Penalties for non-compliance with the reporting regime should be sufficient to*	Civil penalties should apply for failing to monitor or report covered incidents adequately, and for missing timeframes for reporting. Increased civil

	Element & description	An Australian approach
	ensure they are not seen as just a cost of doing business.	and criminal penalties should apply for providing false or misleading information, for repeated breaches, or gross or wilful negligence resulting in non-compliance. Precise penalties will depend upon the nature and circumstance of any breach. However, maximum applicable penalties should be set at a level to ensure obligations carry real weight. For example, under Australia's therapeutic goods regulatory regime, penalties of up to 30,000 penalty units for body corporates and twelve months' imprisonment can apply for failing to notify adverse events within specified timeframes.[4]
	Precedents: *AI Act (EU), art 101 (fines of up to 3% of annual worldwide turnover of €15 million for general-purpose AI providers); Transparency in Frontier Artificial Intelligence Act (California), §22757.15 (civil penalties of up to US$1 million per violation); Responsible AI Safety and Education Act (New York), §1427 (fines of up to US$1 million for first violations, and US$3 million for subsequent violations).*	
6. Implementation		
6.1	**New or amending legislation** *Elements of an AIIMR regime could be implemented through amendments to existing regulatory regimes (such as in relation to data centres under the Security of Critical Infrastructure Act). However, to ensure a cohesive and consistent approach is taken, it is likely preferable for a separate, standalone regime to be established.*	An Australian AIIMR could be instituted via the legislated Australian Standards for AI, if these covered both training and inference, or through separate legislation. Obligations on data centres could be implemented through amendments to the *Security of Critical Infrastructure Act*, although this risks fragmenting reporting obligations under both regimes. Any legislation should establish the framework for the reporting regime but leave specific thresholds and definitions to delegated instruments. This will ensure they can be regularly reviewed to keep pace with technological developments.
	Precedents: *Biosecurity Act 2015, s 154 (definition of 'reportable biosecurity incident' can be set through legislative instrument); Cyber Security (Ransomware Payment Reporting) Rules 2025, rule 6 (sets turnover threshold for reporting entities); Security of Critical Infrastructure Act 2018 (covered assets and other elements set by rules and other instruments).*	

[4] *Therapeutic Goods Act 1989* (Cth), ss 29A, 29AA.