



AI Incident Monitoring & Reporting for Australia

Building trust, reducing systemic risk, and underpinning evidence-based regulation of AI in Australia

The Problem: Government has limited visibility of adverse AI incidents and systemic AI risk in Australia, and avenues for reporting issues are fragmented and unclear

- Artificial intelligence (AI) is rapidly becoming integrated across Australia's economy and broader society.
- The government has limited visibility of AI harms, adverse incidents, and potential systemic failures, hindering its ability to identify and respond to systemic threats, accurately assess risk, and design and implement regulatory responses.
- Disclosure of harm and adversarial misuse by large AI companies has been fragmented, inconsistent, and often dependent upon whistleblowers.
- Businesses and users have limited and unclear options for reporting harm caused by AI, hindering uptake by reducing trust.

The Solution: A national AI Incident Monitoring & Reporting (AIIMR) system

- Australia needs a national system to track and enable responses to significant and systemic AI failures, harms, security incidents, and related near-misses. This includes mandatory reporting by large AI companies and voluntary reporting for other businesses and users.
- Australia already has similar mechanisms for cars, airplanes, medical devices, and consumer goods (see [Box 1](#)), and they are key to ensuring these products are safe. We also have issue-specific mandatory reporting schemes such as for ransomware and privacy breaches.
- Applying a similar system to AI is a low burden, high benefit measure to build trust, reduce risk, and inform regulatory responses.
- AIIMR capability can be scaled from a low cost pilot through to a comprehensive standalone agency (see [Box 2](#)).
- Establishing this capability would also align Australia with leading partners (see [Box 3](#)).

Six Core Functions of an AIIMR Mechanism

1. Surveillance Active monitoring of the AI sector and early detection of potential systemic vulnerabilities.	2. Voluntary reporting An avenue for businesses, researchers, and affected persons to report harm or adverse incidents, particularly where the appropriate regulator for an issue is unclear.	3. Mandatory reporting Obligations on developers and deployers to report adverse incidents and near misses. The scope can scale over time and be aligned with practice such as in the European Union and California, and using OECD guidelines.
4. Coordination Coordinate incident responses across existing regulators; inform future regulation; and ensure cross-cutting consistency.	5. Information sharing Public database of adverse incidents; information sharing within government; and regular reporting on harms and mitigations.	6. Last-resort powers Corrective powers to recall or modify harmful deployments, or enable responses under other legislative frameworks.

For more information about this document, please contact australia@globalshieldpolicy.org.

Box 1. Practices in other regulated industries in Australia

- **General consumer product safety:** The [Australian Consumer Law](#) empowers the Australian Competition and Consumer Commission (ACCC) to enforce safety standards, ban unsafe products, and oversee and [initiate product recalls](#). For example, in 2018 [over four million defective airbags were recalled](#) due to their connection with injuries and fatalities. Where a separate agency is responsible for safety regulation of specific products they are also able to lead a recall when appropriate (including in relation to food, agricultural products, therapeutic goods, and vehicles).
- **Vehicle safety:** Part 3 of the [Road Vehicle Standards Act 2018](#) allows for recalls where a safety issue arises with a vehicle or component that may cause injury or where the vehicle or component does not comply with the applicable standards. The Department of Infrastructure, Transport, Regional Development, Communications, Sports and the Arts has [noted](#) that “*Vehicle and component recalls are a critical part of keeping you and the vehicles on our roads safe.*”
- **Air and transport safety:** The Australian Transport Safety Bureau (ATSB) receives mandatory [notifications](#) of accidents and serious incidents that affect aircraft safety. In 2024, an ATSB investigation into an Airbus EC130 helicopter accident led to the operator enhancing its safety protocols and practices and modifying and upgrading certain equipment. ATSB also made recommendations to the Civil Aviation Safety Authority to improve helicopter safety, which CASA [accepted](#).
- **Medical devices and therapeutic goods:** The [Therapeutic Goods Administration](#) (TGA) has a robust post-market [monitoring](#) and surveillance system for therapeutic goods under the [Therapeutic Goods Act 1989](#). This involves actively monitoring [adverse events](#) and product problems reported by health departments, healthcare professionals, consumers, pharmaceutical companies, and medical device suppliers. The TGA uses this information to identify safety concerns and take regulatory action such as issuing safety alerts, product recalls, or updating product information.

Box 2. Scalable options for establishing AIIMR capabilities

Implementation options range from a low-cost voluntary reporting pilot to a comprehensive mandatory reporting scheme. Based on comparator bodies and overseas experience, we estimate that an AIIMR capability could cost between \$10 million and \$15 million per year. Costs to the Australian taxpayer could be offset by fees on industry.

Baseline Option: Fast, low-cost pilot project

Creation of a function within the Department of Industry, Science and Resources or within the AI Safety and Security Institute to receive voluntary reports from AI companies, broader industry, and users. This option would be faster to implement, likely not require legislation, and could be scaled later. However, it would lack comprehensive coverage without mandatory reporting obligations and have no enforcement powers. It would also trail behind leading jurisdictions.

Key elements:

- Develop incident taxonomies, reporting platforms, and regulatory coordination mechanisms.
- Receive voluntary reporting from AI companies, industry, and users of incidents and near misses, with scope defined through consultation as part of development of taxonomies.
- Undertake analysis and reporting of incidents, near misses, and responses.
- Produce public, anonymised reporting of AI adverse incidents in Australia and measures put in place to respond and mitigate systemic risk.

Mandatory Reporting Option: Mandatory reporting through a dedicated or existing agency

Creation of a standalone agency, or expansion of an existing agency, to receive voluntary reports and mandatory reports from large AI companies and high-risk deployers. This would require new legislation. To minimise the regulatory burden, mandatory reporting obligations should align with other jurisdictions and start with systemic threats, high-risk and catastrophic incidents, and related near misses. Absent specific powers to respond to incidents, the body would rely on other regulators or cooperation from industry for corrective or protective action.

Key elements (in addition to Baseline Options):

- Receive mandatory reporting for a defined set of high-severity incidents from defined entities (e.g. large model developers, critical infrastructure operators, high-risk deployers).
- Support industry preparation for, and response to, incidents alongside other regulators.

Comprehensive Option: Standalone AIIMR agency with corrective powers

A comprehensive approach to AIIMR would require a standalone AIIMR agency with powers to respond to incidents and near misses, including to require industry to take action to protect against or remediate an incident. These powers would be 'last resort' insofar as powers of existing, sector-specific regulators should be the first option when considering a response. This option would require legislation, with ample precedent to draw from in other industries ([Box 1](#)).

Key elements (in addition to Baseline and Mandatory Reporting Options):

- Powers to order corrective action where necessary, such as to recall faulty, malfunctioning or harmful AI deployments, and impose mandatory safety measures to mitigate risk.
- Early-warning functions to alert industry and governments of emerging systemic hazards.
- Integrating and aligning AIIMR functions and powers with existing crisis and emergency response mechanisms, including under the *Security of Critical Infrastructure Act* and the Australian Government Crisis Management Framework.

Box 3. Comparisons of AI mandatory reporting in other jurisdictions

	Body	Covered entities	Scope of mandatory reporting obligation
European Union	EU AI Office and Member State market surveillance authorities.	Providers and deployers of high-risk AI systems (as defined in the AI Act) and certain general purpose AI models.	Serious incidents, including: • death or serious harm; • serious disruption of critical infrastructure; • fundamental rights infringement; • serious harm to property or the environment.
California	Office of Emergency Services.	AI companies that train models using more than 10 ²⁶ FLOPs of computing power..	Critical safety incidents, including: • death or injury resulting from a loss of control of, unauthorised access to, or exfiltration of an AI model; • the materialisation of a catastrophic risk (i.e. death or injury to more than 50 people or more than US\$1 bn of damages arising from CBRN assistance, a lack of human oversight leading to a cyberattack or other crime, or evasion of control); • an AI model subverting control and monitoring in a manner that materially increases catastrophic risk.
New York	New York Department of Financial Services.	AI companies that train models using more than 10 ²⁶ FLOPs of computing power.	Critical safety incidents, including: • death or injury resulting from a loss of control of, unauthorised access to, or exfiltration of an AI model; • the materialisation of a catastrophic risk (i.e. death or injury to more than 50 people or more than US\$1 bn of damages arising from CBRN assistance, a lack of human oversight leading to a cyberattack or other crime, or evasion of control); • an AI model subverting control and monitoring in a manner that materially increases catastrophic risk.
Vietnam	Reporting through national online AI portal.	Developers and suppliers of AI systems.	Serious incidents, namely those that cause or are likely to cause significant damage to life, health, human rights, property, cybersecurity, public order, the environment or disruption to national security information systems.



Annex - Further Resources

[*The California Report on Frontier AI Policy*](#) (Joint California Policy Working Group on AI Frontier Models, June 2025), Section 4 (Adverse Event Reporting).

[*Adverse Event Reporting for AI: Developing the Information Infrastructure Government Needs to Learn and Act*](#) (Issue Brief, HAI Policy & Society Stanford RegLab, July 2025).

[*Safe Beyond Sale: Post-Deployment Monitoring of AI*](#) (Ada Lovelace Institute, June 2024).

[*Message Incoming: Establish an AI Incident Reporting System*](#) (Federation of American Scientists, June 2024).

About Global Shield Australia

Global Shield Australia is an international advocacy organisation dedicated to reducing global catastrophic risk. We advocate for credible and effective regulation of artificial intelligence to reduce its potential for harm and thus ensure that its opportunities can be fully realised.

For more information on this document or our work, please email australia@globalshieldpolicy.org.