Submission to the Productivity Commission's Pillar 3: Harnessing Data and Digital Technology Interim Report Consultation

15 September 2025

About Global Shield Australia

Global Shield Australia is an independent, non-profit policy advocacy organisation dedicated to reducing global catastrophic risk, including from AI. We take an all-hazards approach to global catastrophic risk management, supporting governments to enact and effectively implement policies that prevent and prepare for all forms of risk. We are the Australian office of a global non-profit organisation, which is headquartered in Washington, DC. For more information on this submission please contact australia@globalshieldpolicy.org.

Table of Contents

Executive summary

- **Australia must fully weigh the downside risk associated with advanced AI alongside its potential economic benefits.** AI-enabled harms are already occurring and as models become more capable the potential hazard only increases.
- **Well-designed regulation enables innovation, providing the certainty and consistency needed by industry to invest**. Clear rules foster investment, create a level playing field, and catalyse innovation by reducing legal uncertainty and increasing user trust by delivering safety and security.
- **Seizing the benefits of AI requires uptake, and uptake requires building user trust.** Without public confidence in AI systems, adoption will stall and be uneven. Australians support balanced regulatory responses to AI, that include oversight and safety mechanisms.
- **'Wait and see' and framing AI-specific regulation as a "last resort" are flawed approaches.** A regulatory approach confined to reforming existing frameworks cannot effectively manage the novel and cross-cutting hazards associated with AI. Work on AI-specific regulation must be done concurrently with the regulatory gaps review. Delay will only increase uncertainty, hinder investment, and make regulatory intervention more difficult and expensive in the future.
- **The Government requires better visibility of AI harms.** This can be achieved through a dedicated, economy-wide monitoring and adverse incident reporting mechanism to aggregate data, enable cross-regulator coordination, and support corrective actions.

Global Shield Proposals for Draft Recommendations
Draft recommendation 1.1 Gap analyses of current rules should be expanded and completed as soon as possible.
Draft recommendation 1.2 AI-specification regulation should be included as part of the suite of responses to AI, in particular to target novel threats posed by high-risk and general-purpose AI, address cross-cutting harms most efficiently dealt with upstream in the AI supply chain, and support consistency in regulatory approaches by domain specific regulators.
Draft recommendation 1.3 Continue with the development, enactment and implementation of mandatory guardrails for high-risk and general-purpose AI systems in parallel with the regulatory gaps analysis.
New draft recommendation 1.4 Australia should establish a national AI post-deployment monitoring and reporting scheme to capture adverse incidents and near-miss incidents.

I. Introduction and context

Advanced AI has the potential to transform Australia's economy, with significant productivity gains forecast if adoption is done well. But these benefits are not guaranteed. They depend on Australians—businesses, governments, and the public—having confidence that AI systems are safe, dependable, and governed responsibly.

Trust is the critical enabler. As the Productivity Commission's Interim Report on Harnessing Data and Digital Technology (**Interim Report**) notes, "*consistent and reliable regulation can help promote trust in AI technology, which in turn creates an environment in which business is willing to invest in, develop, and adopt AI*".[1] Yet Australians are among the most cautious globally about AI, and their concerns are justified.[2]

Leading AI developers regularly warn about the extreme harms their own models could cause.[3] And we are already seeing AI being misused for disinformation,[4] harassment,[5] and even in the context of terrorist acts.[6] The impact upon mental health and links to self-harm incidents are also serious causes for concern.[7] There are also clear early warning signs of even greater threats on the horizon, involving relating to bioweapons,[8] cyberattacks, and potential loss of control scenarios.[9]

Further, even a single major AI incident could derail adoption and set back AI adoption by a generation. Other technologies have shown how one catastrophic failure can undermine public confidence for years.[10] Early, proportionate oversight is essential to avoid this outcome for AI.

The Interim Report warns that "*poorly designed regulation could stifle the adoption and development of AI and limit its benefits*".[11] But the Report then underestimates the equal or greater danger of inadequate, ineffectual, or delayed regulation. The Report's downplaying of

---

[1] Productivity Commission, *Harnessing Data and Digital Technology*, Interim Report (August 2025) (**Interim Report**), 10.
[2] Ipsos, *Global Views on A.I. 2023*, July 2023; M. Noetel et al, "80% of Australians think AI risk is a global priority. The government needs to step up", UQ News, 11 March 2024; N. Gillespie et al, *Trust, Attitudes and Use of Artificial Intelligence: A Global Study 2025*, University of Melbourne & KPMG (2025).
[3] Center for AI Safety, *Statement on AI Risk*; Hard Fork, "Dario Amodei, C.E.O. of Anthropic on the Paradoxes of A.I. Safety and Netflix's 'Deep Fake Love'", *New York Times*, 21 July 2023; Sam Altman, "Machine Intelligence, Part 1", 26 February 2015; ControlAI, "What Leaders Say About AI" (2024).
[4] ABC News, "Pro-Russian Influence Operation Targeting Australia in Lead-up to Election with Attempt to 'Poison' AI Chatbots", 3 May 2025.
[5] EducationHQ, "Girls Targeted, Numbers of Deepfake Images Double as Schools Urged to Act", 30 June 2025.
[6] US Department of Justice, "Washington State Man Arrested on Federal Charges Alleging He Provided Material Support to Palm Springs Fertility Clinic Bomber", 4 June 2025.
[7] ABC News, "AI Chatbots Accused of Encouraging Teen Suicide as Experts Sound Alarm", 12 August 2025.
[8] Kyle Hiebert, "AI is Reviving Fears Around Bioterrorism. What's the Real Risk", Centre for International Governance Innovation, 30 June 2025.
[9] Anthropic, System Card: Claude Opus 4 & Claude Sonnet 4 (May 2025).
[10] J. Carlson, "Chernobyl: The Continuing Political Consequences of a Nuclear Accident", *The Interpreter*, 9 July 2019.
[11] Interim Report, 1.

AI's novel and cross-cutting threats[12] misses that existing frameworks cannot, on their own, manage harms embedded at the model layer. Nor is it accurate to suggest that AI-specific regulation is inherently burdensome. Clear, outcomes-based rules can provide certainty, level the playing field, and unlock innovation by reducing legal ambiguity.[13]

The choice for Australia on AI is not between productivity and safety. Productivity relies on trust, and trust requires credible, effective oversight. A robust AI-specific regulatory framework should be seen as the foundation for Australia realising the benefits of AI, rather than a constraint on progress or innovation.

This submission sets out how and why the Interim Report's draft recommendations on AI policy should be strengthened to deliver safe and secure AI, enabling Australia to realise the promised benefits.

Box 1. Addressing Common Misconceptions About AI Regulation

Claim: AI-specific regulation would duplicate existing regulations.

Reality: AI-specific regulation can be designed to avoid duplication and increase consistency and regulatory coherence across other domains. It would act as a baseline safety net for high-risk and general-purpose AI, while leaving domain-specific issues to the relevant regulators. Duplication can be avoided through:

- Recognition of existing regulations where these result in the same desired outcomes; and
- Focusing new requirements on novel, systemic threats, and risk-management practices that cannot be provided for within existing regulatory frameworks.

Claim: AI-specific regulation would be overly broad, capturing too many services.

Reality: Services captured by the proposed AI-specific regulation would by definition be 'high risk'. Regulating AI models would *reduce* the regulatory burden on many applications.

- The Interim Report refers to a potential concern that "*most commercial chatbots*" could be regulated as high-risk AI systems regardless of the efficacy of existing regulators.[14] This concern seems to be premised on the belief that "*most commercial chatbots*" do not pose significant risk. In reality, frontier AI companies' "*commercial chatbots*" are front-ends for highly sophisticated AI models.[15]
- These models require strict security and safety measures to avoid harm, as are currently being voluntarily deployed by some AI companies (with varying degrees of effectiveness). The risk stems from the underlying core models, not just the method in which they are deployed. Regulating these models will encourage uptake and reduce downstream compliance costs.

[12] Interim Report, 15.
[13] Industry has also supported clear regulation overseas, for example, Anthropic, Google, and OpenAI (and others) have all signed the European Union's Code of Practice for General-Purpose AI, see European Commission, The General-Purpose AI Code of Practice, 9 September 2025.
[14] Interim Report, 21.
[15] See, e.g., C. Berg & J. Rosenblatt, "The Monster Inside ChatGPT", *WSJ Opinion*, 26 June 2025.

Claim: AI-specific regulation could "*raise uncertainty*".

Reality: The opposite is true. The absence of clear AI regulation has already created uncertainty for industry. Well-crafted AI-specific regulation can increase certainty by:

- establishing a consistent baseline across sectors;
- minimising fragmented legal disputes on liability for AI systems under various regulatory frameworks; and
- using phased implementation (education and uplift, then enforcement) with rapid review to clarify edge cases and uncertainties.

The Interim Report suggests that new AI-specific regulations "*may*" create uncertainty as they "*may need to be tested before the courts and overlap with other existing regulations*".[16] But this is true for all new regulation, which always *could* lead to legal disputes, as could the status quo.[17] AI-specific regulation is far more certain than leaving firms guessing across multiple potentially inconsistent regimes.

Myth: AI-specific regulation is not "*technology-neutral*".

Reality: Technology-neutral means that the regulation focuses on purposes, outcomes, and risk as opposed to specific technical mechanisms. An AI-specific regulation could be no less technology-neutral than the *Therapeutic Goods Act* or *Civil Aviation Act*, which regulate safety outcomes without necessarily favouring particular technologies. Similarly, an AI regulatory framework can be neutral by referring to functions, roles, and risk, with mechanisms such as incident reporting applying regardless of the algorithm or model type. This approach avoids picking winners and avoids obsolescence as technology develops.

[16] Interim Report, 22.
[17] S. Morrissey et al, "Tech Professionals and AI: Navigating Liability and Emerging Risks", Wotton Kearney, 13 December 2024; D. Perera & A. Ran, "Legal Liability for AI-Drive Decisions - When AI Gets it Wrong, Who Can You Turn To?", HFW, 15 April 2025; K. Frazier, "We're Not Ready for AI Liability", AI Frontiers, 4 June 2025.

II. Why Australia needs AI-specific regulation

Central to the Interim Report's recommendations in relation to AI-specific regulation is the belief that "[f]*ew of AI's risks are wholly new issues*" and that the regulatory focus should be on mitigating AI's contribution of "*additional risk…rather than treating them as wholly new problems.*"[18] This framing, however, mischaracterises the nature of the technology, the risk of harm, and the effectiveness of various regulatory options to respond. Given the significant lack of trust that Australians have of AI, it will be critical to ensure that AI regulation is *credible* and *effective* at preventing potential harms.

A. Advanced AI creates novel and system risk of harm

According to the Interim Report AI-specific regulation is justified where:

(a) Existing regulatory frameworks cannot be sufficiently adapted to handle an issue; and

(b) Technology-neutral regulations are not feasible.

Both of these conditions are already met.

First, in relation to technological-neutrality, as noted above it is a mistake to assume that AI-specific regulation cannot be technology-neutral (see Box 1 above).[19] AI-specific regulation can—and to the extent possible should—be drafted in a way that is technology-neutral, focusing on outcomes rather than prescribing technical methods.

Australia already does this in other domains. The *Therapeutic Goods Act 1989* (Cth) regulates medicines and devices according to their function and therapeutic purpose, not necessarily their chemistry or construction. The *Civil Aviation Act 1988* (Cth) regulates safety and airworthiness across all types of aircraft. The *Telecommunications Act 1997* (Cth) applies to all carriage services, regardless of whether they are delivered via copper, fibre, or satellite. Each of these frameworks has endured and adapted to technological change because they set broad, neutral obligations at the legislative level, while allowing regulators to adapt detailed requirements through subordinate rules and standards.

The same approach can be applied to AI. AI-specification regulation can define obligations by role (such as developer or deployer), by function (such as decision-making or content generation), or by risk category (such as high-risk or general-purpose), without naming specific algorithms or coding techniques. This structure avoids favouring one technological approach over another, reduces the need for constant amendment, and ensures the law remains relevant as technologies develop.[20]

Second, while some AI-related harms are familiar in type, the new ways in which these harms can be realised due to AI are novel in their scale, speed, autonomy, and opacity. This makes

[18] Interim Report, 15 (emphasis removed).
[19] A. Ojanen, "Technology Neutrality as a Way to Future-Proof Regulation: The Case of the Artificial Intelligence Act", *European Journal of Risk Regulation* (2025, online).
[20] L. Moses, "Recurring Dilemmas: The Law's Race to Keep Up With Technological Change" (2007) *University of New South Wales Law Research Series*, No. 2007-21.

existing regulatory frameworks ill-equipped to respond. It also suggests that their ability to properly mitigate the risk will be uncertain, variable, and involve a time of significant confusion as various regulators adopt their own responses to AI developments.

AI also introduces genuinely novel avenues for harm, including through:

1. Deceptive or manipulative model behaviour;[21]
2. Manipulation of AI agents by hostile actors or other AI systems;[22]
3. Self-propagation or 'escape' scenarios leading to loss of control;[23]
4. A lowering of barriers for malicious actors across cyber, biosecurity, disinformation;[24]
5. AI actions and decisions falling through gaps in responsibility, accountability, and liability regimes;[25] and
6. Systemic harm through single, general-purpose models being deployed across multiple domains.

A fragmented, sector-by-sector approach cannot manage these novel threats. Instead, it creates duplication and legal confusion, relying on a reactive regulatory posture instead of considered, proactive measures.

B. Harms are best addressed at their source

The cross-cutting and systemic risk created by advanced AI are most effectively managed upstream, at the point where AI models are built and deployed. Domain specific reforms cannot keep pace with threats and hazards that originate within foundational systems themselves and propagate across multiple domains.

Amending every individual regulatory regime to account for model-layer threats would be inefficient and incomplete compared to a single cross-cutting framework. AI-specific regulation can also mitigate the risk of this harm occurring in the first place, not just prohibit the harm or put in place responses should it occur.

While some AI-enabled threats can be partially mitigated through domain specific reforms—for example, stronger surveillance around biological facilities and DNA printers—the underlying risk vector is the design and release of advanced AI models.

This is equally true for issues associated with deepfakes and synthetic media. Criminal law, electoral law, and media regulation can address specific abuse cases, but none of these regimes

[21] P. Park et al, "AI Deception: A Survey of Examples, Risks, and Potential Solutions" (2024) 5(5) *Patterns*.
[22] K. Greshake et al, "Not What You've Signed Up For: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection", *Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security* (2023), 79.
[23] J. Schlatter et al, Shutdown Resistance in Reasoning Models, *Palisade Research*, 5 July 2025.
[24] Anthropic, *Threat Intelligence Report* (August 2025). .
[25] G. Sadler et al, *AI Legislation: Stress Test*, Good Ancestors (August 2025).

can solve the provenance problem at scale.[26] Conversely, AI-specific regulation can require developers to implement watermarking, metadata standards, or other integrity mechanisms that enable downstream regulators to act effectively.

Direct regulation of AI will also support Australian deployers and businesses integrating AI into their systems and workflows. At present, these businesses and users may be exposed to liability for risk they have little control over or visibility of. Regulating AI model developers can help remedy this imbalance, for example, through certification and disclosure requirements.

There are a range of AI harms that are best prevented and mitigated at their source, through targeted AI-specific regulation that complements sectoral rules rather than replicating them. While it may be possible to address some AI harms by reforms to existing regulatory frameworks, it is not feasible, efficient, or the most appropriate way to do so.

Other cross-sector products and technologies already receive their own specific regulation. For example, industrial chemicals are used in multiple industries but their safety is not left to fragmented sectoral laws; instead, they are governed through a dedicated regulatory framework. AI, with its general-purpose nature and capacity for cascading harm, warrants a similar approach.

A clear regulatory framework that establishes baseline safety and transparency requirements for models is the most coherent way to manage a risk that is not confined to any single domain of use and can only be properly addressed by the model developers themselves. Such a framework could focus on mitigation measures that only model developers can implement—such as model testing and certification, incident reporting, and provenance measures.

C. Monitoring and oversight mechanisms are needed

Given the cross-cutting nature of AI risk, there is also a need for economy-wide monitoring of AI harms alongside incident reporting mechanisms to ensure the government is aware of and therefore able to respond to issues as they arise.

AI risk cuts across industries and regulatory portfolios. A single model can be deployed in health, finance, logistics, and defence simultaneously. If that model has a design flaw, a biased dataset, or a security vulnerability, the resulting failures may appear to individual regulators as isolated problems but in reality reflect a systemic hazard. At present, Australia has no consistent mechanism for detecting, reporting, and responding to this risk.

AI-specific regulation could establish an economy-wide monitoring and reporting mechanism. This would include:

1. Registration of high-risk and general-purpose AI systems;

[26] See work by the Australian Signals Directorate and its international partners on provenance: *Content Credentials: Strengthening Multimedia Integrity in the Generative AI Era*, January 2025, U/OO/109191-25, PP-25-0336.

2. Mandatory reporting of adverse incidents and near misses by industry providers, and voluntary reporting for those users affected by AI incidents; and
3. Ongoing surveillance to detect systemic vulnerabilities and allow for early identification of emerging threats and hazards.

Such a mechanism would give the government the visibility it currently lacks, enable regulators to act in a coordinated and consistent way, and provide Australian deployers and users a clear avenue for reporting and addressing harm. Crucially, it would allow issues to be addressed early, before they spread across multiple domains or cascade into a catastrophic event.

D. Amendments to the Draft Recommendations

In light of the above, we strongly urge the Productivity Commission to amend Draft Recommendations 1.1 and 1.2 to reflect the need for timely and effective AI-specific regulation:

Amended Draft recommendation 1.1
Gap analyses of current rules should be expanded and completed as soon as possible. These reviews should consider current and foreseeable future uses of and harms from AI, identifying where in the AI supply chain regulation would be most efficient.

Amended Draft Recommendation 1.2
AI-specification regulation should be included as part of the suite of responses to AI, in particular to target novel threats posed by high-risk and general-purpose AI, address cross-cutting harms most efficiently dealt with upstream in the AI supply chain, and support consistency in regulatory approaches by domain specific regulators.

AI-specific regulation should be considered as part of the suit of responses to AI. It should:

- Be targeted at use cases and impacts that cannot be sufficiently or efficiently handled by existing regulatory frameworks, with immediate consideration of high-risk and general-purpose AI;
- Be technology-neutral in approach; and
- Support regulatory consistency and coherence with clear definitions, principles, and risk management requirements.

We also propose a new Draft Recommendation 1.4:

New Draft Recommendation 1.4
Australia should establish a national AI post-deployment monitoring and reporting scheme to capture adverse incidents and near-miss incidents.
This should include registration of high-risk and general-purpose AI systems, mandatory and voluntary reporting of AI incidents and near misses, and ongoing surveillance to support coordinated regulator responses.

III. Why 'wait and see' won't work

Delaying AI-specific regulation is not a cautious or prudent strategy, it is a dangerous one. We already have ample evidence of harm from current AI systems, and clear warning signs of how threats will intensify as capabilities advance.[27] This is particularly true in relation to the potential catastrophic harms, which existing regulatory frameworks struggle to address.

Waiting for another cycle of reviews before acting would only heighten uncertainty, slow investment, and risk repeating the mistakes made with social media, where delayed action forced Australia to play catch-up after harms had already become entrenched.

Once AI systems are deeply embedded across the economy, corrective action will be slower, costlier, and less effective. Early action will also ensure regulators build the capability to oversee AI before a serious incident occurs.

Framing AI-specific regulation as a "*last resort*" is particularly problematic.[28] It implies that the government should amend every existing regulatory framework before considering dedicated rules for AI. However, a regulatory gaps analysis will struggle to keep pace with technological development, necessitating concurrent measures to mitigate known and emerging threats.

Concerns about "*burdensome regulation*" stifling innovation or the "*mere threat of regulation*" hindering investment sets up a false choice.[29] Poorly designed or excessive rules can certainly hinder investment, but so can uncertainty. Whether a regulation is excessive will also depend upon the nature and extent of the risk of harm that it is seeking to mitigate.

Currently, the absence of clear and proportionate regulation has already created significant uncertainty for Australian industry.[30] By contrast, a principles-based AI regulatory framework would give businesses clarity, level the playing field for responsible actors, and help unlock AI's productivity benefits.

Australia should not wait for a major AI incident before acting. A proactive, principles-based framework—with mandatory guardrails for high-risk and general-purpose AI and appropriate monitoring mechanisms—is needed now to anticipate and mitigate harm before it materialises.

As such, we urge the Productivity Commission to amend Draft Recommendation 1.3 as follows:

> **Amended Draft recommendation 1.3**
> **Continue the development, enactment and implementation of mandatory guardrails for high-risk and general-purpose AI systems with urgency and in parallel to the regulatory gaps analysis.**

[27] Global Shield Australia, AI Risk Primer: Why Safeguards Are Essential (2025); D. Acemolu, "Harms of AI" in J. Bullock et al (eds), *The Oxford Handbook of AI Governance* (2024).
[28] Interim Report, Draft Recommendation 1.2.
[29] Interim Report, 16.
[30] See Interim Report, 16, where AIIA, MYOB and Montu are all cited as highlighting regulatory uncertainty and the lack of clarity around guidelines and liability frameworks as issues of concern.

IV. Conclusion

Unlocking AI's productivity dividend depends on proactive, well-designed regulation, not a "*wait and see*" approach. Delay invites the very uncertainty that stifles investment and puts Australia on track to repeat past mistakes, where regulatory inaction allowed harm to become deeply embedded before meaningful intervention occurred.

As such, the Productivity Commission's final report on *Harnessing Data and Digital Technology* must properly reflect the credible downside risk associated with AI in its recommendations. The potential threats and hazards are already clear, as is the case for AI-specific regulation.

A balanced approach to establishing Australia's AI regulatory framework—combining accelerated gap reviews, concurrent AI-specific regulation, mandatory guardrails for high-risk and general-purpose AI, and a national incident monitoring scheme— offers the clearest route to delivering safe and secure AI, and thereby enabling Australia to seize AI's benefits.

By acting now, Australia can shape safer pathways for AI development, build the public trust needed for confident adoption, and give businesses the certainty they require to invest. This can help ensure that AI becomes a driver of prosperity for all Australians.