Why Australia Needs an AI Act

Australia's artificial intelligence (AI) policy debate is at a crossroads. Decisions made in the coming months will determine whether Australians are able to trust AI and capture its benefits or whether we are left exposed to unacceptable and avoidable risk.

Calls to wait for yet further reviews ignore the evidence that action is needed now. Australia cannot afford to repeat mistakes made with previous technological developments, such as social media, where delayed action left communities exposed and regulators playing catch-up.

Global Shield Australia has prepared the attached primer setting out *Ten Reasons Australia Needs an AI Act*. Now is the time for the government to lead and deliver the safeguards that Australians expect and need to fully capture the AI dividend.

TEN REASONS AUSTRALIA NEEDS AN AI ACT

An AI Act is the best way to:

1. **Address the unique hazards posed by advanced AI models**, including their demonstrated capabilities to deceive, self-replicate, and pursue their own goals.
2. **Establish monitoring and incident reporting requirements for AI** across all sectors, to track and respond to harms and better understand its systemic risk.
3. **Require AI model developers to take measures to prevent their products causing harm** in all use cases and be transparent to users and regulators.
4. **Enable a consistent and certain approach to AI regulation** across the fragmented regulatory environment that currently applies to its use.
5. **Mandate content provenance and labelling at scale** by regulating generative models, a key tool to deal with harms ranging from deepfakes to disinformation.
6. **Ensure a consistent approach is taken to assigning legal responsibility for AI actions**, including between foreign and local developers, deployers, and end users.
7. **Put in place specific security requirements and standards** to prevent the misuse of advanced AI models by rogue actors seeking to undermine our national security.
8. **Deliver uniform assessment and certification of AI models and tools**, making it easier for small businesses and users to use and trust the technology.
9. **Align Australia with global partners**, minimising compliance friction and enabling Australian businesses to access global AI assurance opportunities.
10. **Ensure dedicated regulatory oversight of AI**, building capable regulators that can set, adapt and implement baseline standards and coordinate across sectors.

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

Ten Reasons Australia Needs an AI Act

Australia's existing regulatory frameworks were developed without AI in mind and cannot, even with amendment, provide the consistent, economy-wide, forward-looking safeguards required for such a transformative technology. A holistic approach to AI regulation is needed to ensure Australia can innovate with confidence while protecting the public from systemic harm. This Global Shield Australia primer sets out why Australia needs an AI Act.

1. Advanced AI models pose unique hazards that are best addressed through an AI Act

Amendments to existing regulatory frameworks can potentially manage specific AI harms or the use of AI in specific sectors. However, they cannot efficiently deal with the novel and cross-cutting hazards posed by advanced AI models (especially as they move towards general intelligence).

These hazards include:

1. **Deception**: AI models deliberately misleading users about the models' intentions or actions, the effect of which compounds if domain-specific regulation assumes a minimum level of interpretability or honesty from an AI model.
2. **Jailbreaking**: users bypassing safeguards to make AI produce harmful outputs. This can occur in any sector - for example, a sales chatbot (subject to the Australian Consumer Law) being jailbroken to produce offensive material.
3. **Hijacking**: AI agents being manipulated by hostile actors when engaging with public material or applications and pursuing instructions contrary to their user - such as to disclose personal or sensitive information.
4. **Self-propagation and escape**: AI models have been shown to have the capability to seek to copy themselves without authorisation, creating potential proliferation and loss control threats.
5. **Autonomous goal-seeking**: systems may resist shutdown attempts or try to pursue objectives against users' or developers' intent.
6. **Training data poisoning**: malicious or flawed training data creating hidden vulnerabilities in the resulting AI model. Without standards or regulation in regard to how training data is collected or cleaned developers risk models being biased or subject to unknown vulnerabilities.

These are not hypothetical concerns. These are hazards that are already being observed in testing of frontier models. No existing regulatory regime is designed to anticipate and mitigate the issues across all domains, and they could manifest in applications in any industry. The most efficient approach, therefore, is to address these hazards at their source—namely during model design, testing, and deployment.

Only an AI Act can ensure these hazards are addressed comprehensively and at the point of greatest impact. An AI Act would address anticipated and unanticipated harms that no amendment to privacy, workplace, consumer, or other more specific regulatory frameworks could cover.

2. Only an AI Act can impose monitoring and incident reporting requirements for AI deployments across all sectors

An AI model or system does not operate or fail in isolation. Because one AI model can be deployed in multiple industries, a single algorithmic flaw, biased dataset, or security vulnerability can cause widespread, systemic harm or simultaneous failure modes across multiple regulatory frameworks. This can result in multiple regulators facing what they see as single failures but missing a potentially systemic hazard. It also means there is no consistent avenue for reporting harm caused by AI deployments.

Australia already recognises that technologies with cross-sector uses can require their own regulatory regime. For example, industrial chemicals are used in many industries, but we do not regulate them solely through sector-specific laws. Instead we have a specific regulatory framework dedicated to industrial chemicals. The same principle should apply for AI models.

An AI Act can mandate monitoring and reporting obligations for developers and deployers of AI economy-wide. This includes measures such as registration of high-risk systems, adverse incident reporting, and on-going surveillance. This would give the government the visibility to respond when systemic issues arise and ensure regulators can act in a coordinated and consistent way across all domains.

3. Only an AI Act can require AI model developers to (a) take measures to *prevent* harm across the spectrum of high-risk applications of their products and (b) disclose the capabilities of their models and safeguards they've put in place

Existing laws, such as for consumer protection, can address harm after it occurs or prohibit certain uses; but they may be less suited to imposing pre-deployment duties on developers at the model level. Ensuring models themselves are safe is key given that defects in a foundational model are inherited across all its downstream applications.

These duties could include requirements to test and certify AI model safety before release and to be transparent with regulators and users regarding how a model was trained. The data and methods used to train advanced AI models, while often proprietary, also strongly shapes their potential for bias, error, and harmful outputs. Without disclosure, regulators cannot properly assess developer claims or the safety of deployed systems.

An AI Act can set clear safeguards for AI models: such as prohibiting unacceptable uses, requiring pre-deployment testing, staged releases, and recall powers. It can also require disclosure to regulators (and where appropriate the public) of information regarding training data, evaluation results, and other key documentation.

4. The current regulatory environment for AI is fragmented, with potential inconsistencies and uncertainties that an AI Act can most efficiently resolve

At present, the same AI model can fall under multiple regulatory frameworks depending on how and where it is deployed. This can create uncertainty for developers, deployers, and users, with compliance obligations potentially duplicative, unclear, and inconsistent. Without a baseline, each regulator can apply different definitions and requirements, meaning the same AI system could be treated differently depending on the sector in which it is deployed. This raises compliance costs, and undermines effective safeguards.

An AI Act can provide a uniform baseline of standards and duties, ensuring regulatory consistency across all sectors. This would reduce opportunities for regulatory arbitrage, and ensure Australians are protected by minimum, clear, and coherent rules no matter where or how AI is deployed. It would also *enable* innovation not chill it. By setting clear limits and obligations, an AI Act would allow safe experimentation and protect responsible innovators from being undercut by unsafe actors.

5. Only an AI Act can mandate content provenance and labelling at scale

The rapid rise of generative AI is making it increasingly difficult to determine whether images, video, or audio are genuine or artificially generated. Existing laws can prohibit harmful content and regulate particular expressions or uses of these tools, but they cannot easily resolve the underlying problem of provenance—knowing what has been created by AI in the first place.

An AI Act can mandate that developers and deployers embed provenance signals such as metadata, watermarking, or equivalent measures at the source. This would give regulators, platforms, and the public the ability to detect AI-generated media at scale, strengthening safeguards against misinformation, fraud, and other misuse.

6. An AI Act can provide the foundation for ensuring a consistent approach to legal responsibility for AI use across the economy

Current regulatory frameworks may struggle to apportion responsibility in complex AI supply chains involving foundation model developers, application developers, and end-user deployers. Upstream developers can also use contractual terms to shift liability, even if they are the ones best placed to address or remedy the harm. For example, if an AI tool used to assess rental applications is found to be racially biased, liability could fall on multiple parties, including the real estate agent, the developer of the assessment application, and the AI company that trained the foundation model.

An AI Act can establish a consistent, economy-wide baseline framework for allocating responsibility for actions by AI tools. It could ensure that upstream developers remain accountable where they are most able to remedy a risk or harm, and provide regulators with a clear baseline to adapt within their specific regimes.

7. An AI Act can put in place specific security standards to prevent the misuse of advanced and high-risk AI models by rogue actors

Frontier AI models are prime targets for intellectual property theft and misuse by criminal groups or hostile State and non-State actors. Without dedicated safeguards, critical components such as model weights, training data, or deployment architectures could be stolen, leaked, or repurposed for malicious use. Existing security regulations are generally not designed to address these risks at the model level.

For example, a single model could raise distinct national security concerns across multiple domains. It could be used to enable biological weapons threats while also contributing to cyber attack capabilities. Without a coherent framework directed at the underlying model, these threats would need to be addressed multiple times under multiple frameworks.

An AI Act can mandate minimum security controls for advanced and high-risk models, ensuring protections are in place before deployment. It can also establish a clear taxonomy of "high-risk" and "nationally significant" AI systems, providing a foundation for consistent obligations across other regulatory regimes.

8. An AI Act can deliver uniform assessment and certification of AI models and tools

Businesses and consumers need confidence that AI systems meet consistent safety and security standards. Without a uniform framework, industry and consumers risk confusion, fragmented certifications, and "safety washing" by companies making unverified claims.

An AI Act can establish clear conformity assessment processes and trusted certification schemes—such as an AI Safety or Trust Mark—that apply across all sectors and supply chains. This would provide a single, recognisable signal of compliance, making it easier for small businesses to adopt AI safely and for users to trust the technology.

9. An AI Act would align Australia with global partners

While industry may oppose regulation as potentially "scaring away" leading AI developers, appropriately drafted AI legislation would align Australia with overseas jurisdictions such as the European Union, and with OECD and G7 principles. Even in the United States, recent efforts to ban state-led legislation that provides robust regulation (such as in California) were broadly rejected by Congress with a vote of 99 to 1 in the US Senate. As such, most major AI companies will already need to comply with overseas requirements.

Without an AI Act, Australia risks becoming a mere rule-taker or jurisdiction of convenience. By aligning with international standards, an AI Act would reduce compliance friction, give Australian firms access to global assurance ecosystems, and ensure our businesses can compete and collaborate on a level playing field.

10. An AI Act can ensure dedicated regulatory oversight for AI in Australia

AI cuts across every sector of the economy, yet no existing regulator has the mandate, expertise, or powers to oversee it systemically. Today, responsibility is fragmented across privacy, consumer, financial, health, and competition regulators. Promises of the potential of AI are also premised on it being a truly revolutionary technology. This means it will need dedicated oversight to manage the profound changes it will bring.

An AI Act can ensure there is dedicated oversight of AI in Australia, with clear authority to set and enforce baseline standards, develop regulatory capabilities, coordinate across regulators, and update regulatory requirements as technology evolves.

About Global Shield Australia

Global Shield Australia is an independent, non-profit organization dedicated to reducing global catastrophic risk. We advocate for credible and effective regulation to minimise AI risk and maximise its benefits. For more information on this primer or our work, please contact australia@globalshieldpolicy.org.