aiia
australian information
industry association

Australian Information Industry Association

Submission on

Proposals for Mandatory Guardrails for High-Risk AI Settings

4 October 2024

aiia
australian information
industry association

Introduction

The Australian Information Industry Association (AIIA) welcomes the opportunity to contribute to the Department of Industry, Science and Resources regarding the proposed Mandatory Guardrails for AI in High-Risk Settings ('Guardrails'). We appreciate the Department's efforts in engaging with industry stakeholders to develop a framework that ensures the safe and responsible development and deployment of Artificial Intelligence (AI) technologies in Australia.

Our position outlines support for clear, risk-based regulations while advocating for a framework that remains flexible enough to foster innovation and growth within the AI industry. We prioritise focusing on genuine high-risk uses to ensure a balanced and proportionate approach. Additionally, we address concerns regarding the significant compliance burden on businesses, particularly relating to transparency, accountability frameworks, and global interoperability.

Part 1: Definition of High-Risk AI

We believe that the proposed principles are currently too ambiguous to adequately reflect an appropriate threshold or measure of high risk. If enacted as is, these principles could negatively impact businesses by failing to provide the certainty and clarity needed for the development and adoption of AI technology within appropriate regulatory frameworks in Australia.

For example, there is no clarity on whether a use case is deemed high risk if one, several, or all of the principles are breached—that is, should they be read conjunctively ('and') or disjunctively ('or'). If it is the latter, the broad scope of principles (a) and (e) in the proposals paper could inadvertently capture use cases where the risks are actually low. Furthermore, principle (f) lacks specificity regarding the appropriate measure of 'severity and extent of adverse impact'. We strongly suggest that the measure of 'high risk' should be set at a distinctively high bar, guided by the reasonable test of imminent harm.

For this reason, we recommend implementing a hybrid approach that combines principles-based guidance with a limited list of well-defined prohibited or restricted use cases in defining high-risk AI. This list would provide specific examples of what constitutes high-risk applications, offering greater clarity for businesses and informing any centralised guidance material that the government provides under existing regulatory frameworks. This hybrid model provides essential flexibility while fostering consumer trust in the technology and maintaining necessary protections for the public and the economy.

However, we urge caution in expanding this list too broadly. The framework must allow for innovation and flexible application of AI technologies without overly restrictive boundaries. A scheduled review process for the use-case list would ensure it remains relevant as technology evolves.

Regarding the use-case list specifically:

1. **Emphasise Imminent Harm**: The definition should emphasise principle (b) such that an AI system is only deemed high risk when there is a clear and imminent threat of harm.

2. **Narrow Focus**: A list of high-risk use cases should remain narrowly focused on areas where AI applications present clear risks to human rights, privacy, or safety.

3. **Regular Review**: This list should be limited and reviewed regularly to ensure it remains relevant. Adaptive legislation, such as delegated legislation, could be used to allow revisions to regulatory frameworks based on advancements in AI and data.

4. **Mechanism for Disagreements**: There should be a clear mechanism to address potential disagreements, e.g., an AI deployment that falls under the high-risk list but is not perceived as high risk by the developers.

Blacklist

At present, the AIIA agrees with the following being used as examples of high-risk use cases:

1. **Exploitation of Vulnerabilities**: Exploitation of vulnerabilities of persons, manipulation, and use of subliminal techniques.

2. **Social Scoring to Refuse Service**: Social scoring to unilaterally deny or allow access to public and private services.

3. **Predictive Policing to Restrict Freedoms**: Individual predictive policing based solely on profiling people.

4. **Emotion Recognition to Mete Punishments**: Emotion recognition in the workplace and educational institutions for strictly punitive purposes.

Regulation of General-Purpose AI (GPAI)

We do not support the default classification of all General-Purpose AI (GPAI) as high risk. The hybrid approach for defining high-risk AI proposed in this submission is flexible enough to capture high-risk use cases without the need for blanket categorisation.

The assumption that GPAI is inherently high risk due to its nature and capabilities is not accurate and could inadvertently capture low risk uses. There are many examples of potentially low risk use cases of GPAI, including customer service chatbots and small language models employed in non-critical applications. Over-regulating GPAI in its nascent stage could stifle innovation and impose requirements that may not align with the actual risks posed by specific uses.

Whitelist

The AIIA also suggests that the Department consider a whitelist for AI use cases that demonstrably serve the public good. One example is to exclude AI systems used for cybersecurity purposes from being classified as high risk. AI is central to defenders' ability to protect our digital way of life from increasingly sophisticated cyber threats and should be leveraged for cybersecurity defence. Policymakers should carefully consider the varied nature of AI use cases to ensure that any new guardrails do not unintentionally inhibit the continued and expanded use of AI-powered tools for good such as cyber defence. This approach would align with principles recognised in both the Colorado Artificial Intelligence Act and Recital 55 of the EU AI Act.

Part 2: Proposed Mandatory Guardrails

A key concern is the potential burden on businesses that may lack the resources to fully comply with complex regulations. Consistent education, clear guidance, and support should be provided to ensure organisations can understand and implement the Guardrails effectively without requiring costly legal or compliance assistance. The Guardrails should prioritise practical, scalable enforcement that is feasible for organisations across varying capacities. It is crucial to strike a balance in the complexity of compliance to avoid hindering innovation, especially among smaller companies driving advancements in AI development.

Global Interoperability

We emphasise the need for global interoperability in conformity assessments, particularly under Guardrail 10. Aligning Australia's assessment processes with those of key international jurisdictions, such as the EU, Canada, and standards like ISO's AI Management System 42001 and the NIST AI Risk Management Framework, would significantly reduce the regulatory burden on businesses and future-proof our approach. Streamlining conformity assessments in this way would not only save resources but also encourage smoother global operations, supporting innovation without compromising regulatory goals.

Third-Party Audits

Another concern relates to the implementation of Guardrails 4 and 10, specifically regarding the testing and validation requirements. We seek to ensure that third-party audits are not mandated, allowing companies to demonstrate compliance through robust internal processes provided they have the requisite capacity to meet regulatory standards.

This is driven by the current lack of robust infrastructure for third-party independent audits in AI and machine learning within Australia or beyond—including the absence of consensus on professional standards for AI system auditors, and the lack of a governing body to establish baseline standards or enforce professional ethical frameworks for their operation.

As the proposals paper acknowledges, there will be a need to consider options in the short versus long term, recognising that the assessment and certification infrastructure for AI standards are still developing and will require time to mature, meaning that insisting on untested and prescriptive requirements for third-party conformity assessments could create material barriers to deployment of AI systems.

Supply Chain Transparency - Distinguishing Developers from Deployers

The AIIA believes that the proposed definitions of 'developer' and 'deployer' are insufficiently distinctive and may cause confusion due to overlap and potential scope creep, especially amongst Application Programming Interfaces (APIs).

Current proposed definitions with AIIA emphasis in bold

Developer: organisations or individuals who design, build, **train, adapt**, or combine AI models and applications.

Deployer: any individual or organisation that supplies or **uses** an AI system to provide a product or service. Deployment can be for internal purposes or used externally impacting others, such as customers or individuals

For example, the proposed definition of 'deployer' is broad enough to encompass end-users. Additionally, because AI is iterative and continually learning, a deployer whose outputs are used as inputs to continuously train AI models risks being regarded also as a developer.

There is a need to improve the distinction between developers and deployers to ensure a clear distribution of liabilities across every part of the complex AI supply chain. We emphasise the need for a more tailored approach, suggesting that procurement and contractual obligations would be more effective in the short term in managing responsibilities throughout the supply chain, rather than relying on desk-based policies.

We recommend that the Department conduct an exercise to map the risk levels and existing developers and deployers against the supply chain to simulate the effectiveness of the current proposals. We also suggest establishing an industry sandbox to learn from real-world applications and to refine the regulatory framework. With these lessons, eventually, a principle of shared responsibility could be developed for implementing guardrails across the supply chain that reflects that there are different obligations for different activities based on the AI technology architecture (e.g., for developers of AI models, developers of AI applications, and deployers of AI applications).

There are significant challenges in applying regulations to supply chains and the complex real-world ecosystem, highlighting the lack of clarity in proposed obligations and the complexity of different industries' supply chains. We stress the importance of avoiding a rush into legislation without proper implementation and learning. Establishing a framework to manage misalignments and misinterpretations is crucial.

Managing Compliance Burdens on SMEs
While transparency across the AI supply chain is critical, particularly to ensure the safety and ethical use of AI systems, the implementation of these requirements must be practical and scalable. Smaller businesses may not have the resources to fully meet complex transparency obligations, especially if these involve extensive documentation or third-party verification. We recommend that the Guardrails clarify the specific obligations at different points of the supply chain, ensuring that accountability is appropriately distributed.

Regulatory Sandboxes
Additionally, we propose the introduction of regulatory sandboxes to ensure small businesses can innovate with AI without being overwhelmed by regulatory burdens. These sandboxes provide an environment where businesses can test AI systems in real-world scenarios without needing to meet full regulatory compliance immediately.

There is also a growing prevalence of these programmes globally, with the OECD AI Principle 2.3 recommending that governments ‘consider using experimentation to provide a controlled environment in which AI systems can be tested and scaled up as appropriate.’ Additionally, the EU AI Act Article 57 mandates: ‘Member States shall ensure that their competent authorities establish at least one AI regulatory sandbox at national level, which shall be operational by 2 August 2026.’

The use of regulatory sandboxes offers the government a strategic tool to gather valuable data on the impacts of AI and thereby contribute to the refinement of regulatory frameworks. Singapore's AI Verify model is a good example of this. This dynamic approach encourages collaboration between the government and AI players to ensure that risks are addressed proactively while also promoting an environment where safe experimentation is encouraged. This public-private collaboration is an important strength as the pace and complexity of AI development increases.

Data Disclosure Requirements
We understand in Guardrails 3 and 6 the government's intention to encourage data transparency as a way to determine a model's performance for a particular use case. However, we raise concerns that requiring businesses to disclose training data sources, datasets, and collection processes would result in industry having to share highly commercially confidential information with significant competitive value—which would be highly inappropriate and at odds with trade secrets law.

Furthermore, and more to the government's direct objectives, we raise concerns that requiring visibility into training datasets will not actually assist deployers in making the determination as to whether a model's performance is appropriate, accurate, unbiased, and fair. Simply because a model has been trained on certain data does not mean it will perform as needed for a deployer's specific use case. Instead, we encourage the government to test a model's outputs, which is the industry-standard mechanism to provide the best indication of performance, and is the best way to foster public confidence and consumer trust in a model.

Part 3: Implementation Framework

Option 3 - Incompatibility of Adopting a Monolithic Legislation (e.g., the *EU AI Act*) in the Australian Context

The AIIA has consistently disagreed with Option 3, which the industry has fed back as too heavy-handed and could stifle innovation. The one-size-fits-all prescriptive law ignores the fast pace of AI development and complexities of the digital economy and multi-layer regulatory framework. In particular, it could be seen as overriding existing regulations, including the already updated *Criminal Act* to address deepfake pornography and the proposed changes to update the *Privacy Act* to increase transparency in the use of automated decision-making.

Furthermore, Option 3 is not technology-agnostic and risks the application of new technology-specific legislation that has economy-wide impact, even where there may be only a small regulatory gap (concerned with unknown future risks), such as a *Quantum Act* or *Biomedical Implant and Devices Act*. The *EU AI Act*'s prescriptive approach risks becoming unresponsive to rapid technological changes, as demonstrated by the following two examples.

Firstly, despite clarifications, its prescriptive definition of AI locks the legislation to regulate general software like an Excel auto-sum function, due to ambiguous terms such as 'inference' and 'autonomy'. Secondly, the Act sets a risk threshold based on computational power (10^{25} FLOPs), conflating computational capacity with risk—a standard that may soon become obsolete due to technological advancements. These examples indicate that the Act will require significant revision due to the ubiquity of embedded AI, swift technological evolution, and changing cultural norms around acceptable AI use.

The AIIA notes too that the *EU AI Act*, when paired with delegated legislation, gives rise to concerning legal principles which can lead to substantial compliance costs. Alongside the EU AI Act, the EU Product Liability Framework reforms will be expanded to include AI systems for the first time and reverse the burden of proof in certain circumstances, so claimants no longer need to prove elements of their case. In particular, the revised regime introduces circumstances in which defect or causation can be presumed.[1]

Two of these circumstances are:

1. Non-Compliance with Regulations: Where there is non-compliance with relevant EU product safety regulations.

2. Technical Complexity: If it is excessively difficult on account of the technical or scientific complexity of a product for a claimant to prove either that:

   - A product is defective; or
   - There is a causal connection between the defect and the damage.

According to the American Chamber of Commerce to the European Union (AmCham EU), while the proposal does not intend to reverse the burden of proof, the presumption of defectiveness and causality effectively amount to a reversal of the burden of proof for products that are particularly technically or scientifically complex. Together, the EU AI Act and the proposed EU Product Liability Framework will create unintended consequences, increasing risks of lawsuits and dampening AI innovation.

Preference for a Narrowly Scoped Option 2

We support the adoption of a unified framework for implementing the Guardrails as suggested in Option 2, as it offers the most effective way to establish a targeted, adaptable regulatory environment that can evolve alongside advancements in AI technologies. This framework allows for overarching principles that apply across sectors while enabling context-specific regulation tailored to the varying risk levels in different industries.

However, rather than explicit legislation, this framework should be principles- and outcomes-based, with the implementation detail left to relevant standards, rather than legislating prescriptive requirements that may quickly become outdated.

The AIIA strongly cautions against scope creep such that Option 2 is effectively Option 3. Instead, the framework should be narrowly focused on identified gaps:

1. **Testing, Transparency, and Accountability Requirements:** Implement requirements that ensure AI systems are appropriately tested and transparent, with clear accountability mechanisms.

2. **Data Management and Governance:** Advocate for a balanced approach to data management and governance for safe innovation rather than disjointed treatment of data across various legislation or programmes (e.g., *Privacy Act, impending Cyber Security Bill, Consumer Data Right, Treasury Laws Amendment (Consumer Data Right) Act, Copyright Act*, etc.).

[1] EU Briefing, EU Legislation In Progress: New Product Liability Directive, see page 6 and 8.

3. **Effective Regulatory Body:** Enact an effective regulatory body or mechanism which must be able to balance the twin national goals of innovation for the good of the digital economy and consumer safety.

4. **Industry Consultation:** Ensure this body consults with and is properly informed by the industry of key developments in AI and its impact in reducing or increasing risks of harm.

5. **Context-Specific Framework:** Develop a framework that is context-specific, noting differing levels of risks in various industry sectors.

6. **Consider Economic and Societal Contexts:** This framework also needs to consider unique economic and societal contexts (e.g., burden on smaller players and cultural considerations). For example, there could be a less stifling compliance system to allow smaller tech firms to comply initially, with the option to scale compliance as they grow. This ensures that innovation from smaller AI players isn't stifled due to high upfront costs.

7. **Government Support for SMEs:** Provide government incentives or support for SMEs to comply with AI guardrails. This includes funding for AI testing and auditing, legal assistance programmes, or tax relief tied to AI safety and transparency compliance efforts.

By being meaningfully narrow, this framework will be set up for success because it is not overly complex, which could intimidate innovative AI companies from entering the Australian market with useful AI products and services and be too difficult to follow by smaller companies.

Conclusion

In conclusion, the AIIA appreciates the Department's efforts in considering the industry's perspectives on the proposed Mandatory Guardrails for AI in High-Risk Settings. We believe that a harm-specific, and balanced, flexible, and proportionate approach is essential to ensure both the safe deployment of AI technologies and the continued growth and innovation within the sector. We look forward to continued collaboration with the Department to develop a regulatory framework that achieves these objectives.

Should you require further information, please contact Ms Siew Lee Seow, General Manager, Policy and Media, at siewlee@aiia.com.au or 0435 620 406 or Mr David Makaryan, Advisor, Policy and Media at david@aiia.com.au.

Thank you for considering our submission.

Yours sincerely
Simon Bush
CEO, AIIA

aiia
australian information
industry association

About the AIIA

The AIIA is Australia's peak representative body and advocacy group for those in the digital ecosystem. Since 1978, the AIIA has pursued activities to stimulate and grow the digital ecosystem, to create a favourable business environment for our members and to contribute to Australia's economic prosperity.

We are a not-for-profit organisation to benefit members, which represents around 90% of the over one million employed in the technology sector in Australia. We are unique in that we represent the diversity of the technology ecosystem from small and medium businesses, start-ups, universities, and digital incubators through to large Australian companies, multinational software and hardware companies, data centres, telecommunications companies and technology consulting companies.