

4 October 2024

Department of Industry, Science and Resources
GPO Box 2013
CANBERRA ACT 2601

Via email: aiconsultation@industry.gov.au

Dear Department

Safe and Responsible AI in Australia – Proposals paper for introducing mandatory guardrails for AI in high-risk settings

Thank you for the opportunity to provide a submission to the Department of Industry, Science and Resource's consultation on the introduction of mandatory guardrails for AI in high-risk settings (**Guardrails**).

The AICD's mission is to be the independent and trusted voice of governance, building the capability of a community of leaders for the benefit of society. The AICD's membership of more than 53,000 reflects the diversity of Australia's director community, comprised of directors and leaders of not-for-profits, large and small businesses and the government sector.

The AICD recognises the need to support directors in implementing effective, safe and responsible Artificial Intelligence (AI) governance. Our program includes a [Director's guide to AI Governance resource suite](#), [webinar program](#), [articles](#) (including highlighting the Voluntary AI Safety Standard to directors) and an AI fluency for directors [short course](#).

1. Executive Summary

The AICD recognises the significant opportunities of AI and the need to incentivise its development and use to remain competitive in the global market and boost national productivity. However, we agree that AI's far-reaching impact and the presence of unique risks require careful management. In summary, the AICD makes the following key points:

1. The scale and breadth of AI and the presence of explainability, bias, and hallucination issues within AI systems can lead to harmful outcomes. We agree that where there is a gap in existing laws, AI use that risks causing such harms to end-users needs to be subject to new regulation.
2. Directors are responsible for the oversight of the organisation's strategy and risk management processes in line with their directors' duties. This includes managing AI risks and opportunities. To facilitate this, directors have an interest in the development of AI regulation that is clear, proportionate and consistent with other intersecting privacy, data and cyber security obligations.
3. The Guardrails need to further consider the different roles and responsibilities of developers and deployers, noting that the majority of Australian organisations are likely to be deployers and/or developers of AI *applications* (as distinct from developers of AI *models*). Issues to be addressed include deployer access to, and reliance on, information on AI system design, training and testing. The threshold at which a 'deployer' becomes a 'developer' (i.e. how much

adaptation of an AI system is required for a deployer to become a developer) should also be further considered.

4. Broadly, we agree with a principles-based, rather than an- EU AI Act style 'list-based' approach to the definition of 'high-risk' AI. However, we are concerned that the proposed principles are too broad and may inadvertently capture low-risk AI uses. Guidance setting out both high-risk and low-risk AI uses may provide greater certainty.
5. We broadly agree with the content of the Guardrails, noting that they appear consistent with other relevant Australian and international principles/frameworks such as the Australian AI Ethics Principles. Interoperability with key developer jurisdictions, including key Australian trading partners, is critical to preserve Australian competitiveness. Given the nascent nature of AI governance and safety solutions, proportionality relief mechanisms, particularly for deployers, may need to be built into the Guardrails (at least initially).
6. Of the three regulatory implementation options, we prefer Option 2 (A framework approach). Given the early stage of AI governance and issues raised with the EU AI Act's strict regulatory approach, a standalone AI Act for the entire economy (Option 3) is not supported and could have unintended negative consequences. Coordination and alignment with concurrent reforms to privacy, data governance and cyber security should be prioritised.
7. Regulation alone is not sufficient to achieve the policy objective of maximising AI benefits while minimising harms. Regulation must be accompanied by actions aimed at uplifting AI capability and governance skills and encouraging innovation. The AICD is committed to lifting awareness, education and competency on AI governance at the director and board level.

2. 'Developer' and 'Deployer'

Relevant consultation paper questions: 10, 12

We agree that requirements should be allocated to the party best equipped to address risks, having regard to access to information (such as AI design and training data) and the ability to make changes to the AI system. We are pleased to see Attachment E of the Consultation Paper sets out key differences between how developers and deployers may satisfy the Guardrails.

However, we highlight four issues:

1. **Guardrails assume ready access to developer information by deployers:** The Guardrails appear to assume that deployers have ready access to information on AI models (including training data, design and testing) from developers, and that this information is accurate and verifiable. The reality is that a significant portion (potentially the majority) of AI *model* developers will be located overseas. Unless developers are either required to disclose this information (which is currently limited to those covered by jurisdictions with mandatory AI disclosure requirements such as the EU AI Act and Colorado) or choose to voluntarily disclose, it may be difficult for Australian deployers or *application* developers, particularly those with less bargaining power such as SMEs, to obtain the necessary information to meet the Guardrails. It is also unclear to what extent deployers are legally able to rely on developer information and assurances on AI system design and testing, and how this impacts potential liability.
2. **Application to international developers and deployers should be confirmed:** It would be useful for the Government to confirm that the Guardrails are intended to capture international developers or deployers where end-users are located in Australia. This is important to clarify, given a substantial number (or even a majority) of Australian organisations are likely to be deployers of AI systems developed by overseas developers.
3. **'Deployer' captures internal use of AI:** The proposed definition of 'deployer' captures internal use of AI system, such as the use of generative AI to improve productivity. This means

organisations purchasing an 'off the shelf' AI product from an international AI developer could be required to apply the Guardrails where such use falls within the broad 'high risk' definition. For instance, the internal use of generative AI by a professional service provider (lawyer or accountant) in the course of providing financial or legal advice may fall within the current definition of "deployer." The need to avoid harm to end-users in such contexts should be carefully balanced with the significant productivity opportunities arising from the use of AI within internal business processes¹ (noting that Australia's productivity is lagging).

4. **'Developer' may capture some deployers:** We are concerned that the current definition of 'developer' is too broad. In particular, the inclusion of 'adapting' an AI model may capture those who make only minor modifications or adaptations to an existing AI system. Given (as we understand it) the intention is to attribute greater legal responsibility to developers rather than deployers, the threshold for shifting from a 'deployer' to a 'developer' should be suitably high. Consideration should be given to the Colorado AI Act's definition of developer, which requires the "intentional or substantial modification"² of an AI system, which is a higher threshold than what the Guardrails currently propose.

3. Definition of high-risk AI

Relevant consultation paper questions: 1 and 3

Broadly, we agree with a principles-based, rather than an- EU AI Act style 'list-based,' approach to the definition of 'high-risk.' This avoids being overly prescriptive (such as classifying whole sectors as 'high risk') and risking low-risk use cases being unintentionally captured. A 'list-based' approach also risks becoming obsolete as AI use cases continue to develop.

However, while we appreciate the principles were intentionally drafted at a high level to preserve flexibility, their breadth and subjectivity lead to a number of potential issues, including:

- **May inadvertently capture low-risk AI uses:** The breadth of the principles may inadvertently capture low-risk AI uses, such as where organisations use AI internally to increase productivity, improve accuracy or reduce inefficiencies. To avoid capturing low-risk AI uses, we would support consideration of a similar 'carve-out' as currently applies in the EU AI Act under Chapter III, Section 1, Article 6(3).³ It might also be worth adopting the approach taken by the Colorado AI Act, which confines 'high-risk' AI use only where "AI is a substantial factor⁴ in making a consequential decision." The inadvertent capture of low-risk AI uses may also be mitigated through the issue of guidance illustrating application across a range of sectors, including examples of low-risk v high-risk AI use within the same sector (the Voluntary AI Safety Standards provide some of this guidance, but we consider this could be incorporated and further expanded in the Guardrails).
- **Self-assessment lacks certainty:** Given the breadth of the principles, there is some concern that a self-assessment process, particularly in the absence of verification or assurance, may lack certainty for businesses. For instance, organisations which are unsure about whether their

¹ [Tech Council of Australia research \(June 2024\)](#) estimates that greater adoption of AI could contribute up to \$115 billion to the Australian economy, with 70% of this to come from internal process-driven productivity gains.

² Which is defined as "a deliberate change made to an AI system that results in any new reasonably foreseeable risk of algorithmic discrimination" – see the [Colorado AI Act](#).

³ Where (i) the AI system is intended to perform a narrow procedural task; (ii) the AI system is intended to improve the result of a previously completed human activity; (iii) the AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment, without proper human review; or (iv) the AI system is intended to perform a preparatory task to an assessment relevant for the purposes of the use cases [designated as high-risk in the Act].

⁴ 'Substantial factor' defined as "(1) A factor that assists in making a consequential decisions, is capable of altering the outcome of a consequential decision and is generated by an AI system; and (2) any use of an AI system to generate any content, decision, prediction or recommendation concerning a consumer that is used as a basis to make a consequential decision concerning the consumer."

AI system or product is 'high risk' and whether they have effectively implemented the Guardrails may spend considerable time and cost in implementation without obtaining the certainty that their AI use does not cause regulatory concern. To aid business certainty, it may be beneficial to introduce a mechanism by which entities seeking to use AI in potentially high-risk context could voluntarily seek confirmation/ approval from relevant authorities. Such a pre-vetting process would likely encourage greater AI use and development, as it would give relevant organisational decision-makers greater comfort that potential harms had been mitigated to the regulator's satisfaction. Such a pre-vetting/ pre-approval process may also need to be mandatory for certain high-risk AI uses which pose a serious risk to national security or critical infrastructure.⁵

- **Breadth of principles will likely increase the compliance burden and cost:** Because of their breadth, some of the principles will require significant analysis and expert input, which would increase the compliance burden and cost. For instance, it would be difficult for an organisation to determine whether an AI use case triggers any of the seven international human rights treaties and six optional protocols Australia is party to, and their legal effect, without legal advice. As noted in the Consultation Paper, Australia does not have a singular, centralised domestic human rights charter or legislation. Further, the process of incorporating international treaties into Australian domestic law is complex and nuanced (that is, the mere fact that Australia is a party to a treaty does not mean it has legally binding status). As such, we caution against taking a broad-brush approach to the incorporation of international human rights treaties. Broad concepts such as 'health and safety,' 'legal effects,' 'impacts to groups or collective rights of cultural groups,' and 'impacts to the broader Australian economy, society, environmental and rule of law' are similarly likely to present implementation challenges.

Other issues which will need to be clarified in relation to the practical application of the high-risk assessment process include:

- The point in time and frequency at which an organisation is required to undertake risk assessments, noting that use cases and risk profiles may shift as the AI system is developed, tested and refined. The incorporation of safe and responsible AI design principles or risk mitigation processes may also itself reduce the risk, severity and extent of adverse impacts;
- Whether a "high risk" rating is necessitated by meeting just one of the six principles, or whether multiple principles need to be triggered;
- Making more explicit that element (f) of the principles ('the severity and extent of those adverse impacts') incorporates an assessment of the likelihood and/or foreseeability of the risk;
- Which person or body within the organisation will be required to make the risk assessment, including confirmation of the nature and extent of board involvement;
- Whether there is any declaration/ sign-off on the risk assessment is required; and
- Whether the risk assessment needs to be disclosed, and if so, what aspects of the assessments need to be disclosed (the final assessment only, or details of the analysis), to whom disclosure should occur (i.e. which government entity or regulator), and where this disclosure will be housed or located (and whether this will be publicly accessible).

⁵ Note that [President Biden's Executive Order on AI](#) requires that companies developing any foundation model that poses a serious risk to national security, national economic security, or national public health and safety notify the federal government when training the model and requires that they share the results of all red-team safety tests.

4. Content of mandatory guardrails

Relevant consultation paper questions: 8

We broadly agree with the content of the Guardrails, noting that they are generally consistent with other relevant Australian and international principles/frameworks such as the Australian AI Ethics Principles.

Interoperability, particularly with jurisdictions developing AI systems (on whom Australian deployers will then rely) and key Australian trading partners, will be critical to harnessing the opportunities of AI, including ensuring that Australia remains an attractive and competitive market.

Some further observations on the content and implementation of the Guardrails include:

- **Need to clarify developer v developer obligations:** Noting our observations in section 2 above, there needs to be clarity on developer v. deployer obligations in respect of each of the Guardrails.
- **Feedback and review mechanism needed:** Given the nascent nature of AI governance and rapid advancement of technology, the regime should incorporate a feedback or review mechanism to ensure the regulation remains fit-for-purpose. It would be prudent to monitor how organisations are implementing the Voluntary AI Safety Standard and incorporate feedback and learnings into the Guardrails.
- **Proportionality relief mechanisms needed:** As the Consultation Paper acknowledges, many of the methodologies and tools needed to give effect to safe and responsible AI are still in development. For instance, AI-generated content labelling and watermarking are still in their infancy, as are methodologies to document data provenance. Further, as noted in section 2 above, in executing some of the Guardrails, deployers are dependent on actions taken and/or information provided by developers. The extent to which deployers can and should be required to verify the quality, completeness and reliability of information received from developers is also currently unclear. In light of these issues, we would encourage the inclusion of relevant proportionality relief mechanisms within the Guardrails. We note that Guardrail 6 does so by proposing a 'best efforts approach', such that organisations are only required to apply the *"best methods available according to their own assessments and with reference to relevant standards."* We consider that such proportionality mechanisms should be applicable to other Guardrails where there is a high degree of uncertainty and/or dependency on third-parties. One potential formulation may be to make guardrails subject to a requirement that entities apply a 'best efforts approach' based on *"all reasonable and supportable information that is available to the entity without undue cost or effort."*

5. Regulatory implementation options

Relevant consultation paper questions: 13 and 16

Option 2 is preferred

Of the three options, we consider that Option 2 is the most practical and balanced.

While Option 1 allows for the greatest integration of AI within existing legislation, we are concerned it would create and exacerbate gaps and inconsistencies in how AI systems are regulated across the economy. Given the breadth of AI application (it permeates all sectors), we are also concerned about the length of time it would take to incorporate AI into each applicable piece of legislation, and the cumulative regulatory burden.

We do not support option 3 at this early stage. AI governance and regulation globally is relatively nascent, with only the EU imposing prescriptive requirements. The EU's approach has already prompted concerns over practical implementation (including lack of certainty caused by broad definitions), the stifling of innovation and whether the regulation is suitably 'future proof.' Given Australia's role in the global AI value chain, we are concerned about moving to a whole-of-economy AI regulation before we can benefit from the learnings from leading developer jurisdictions. The creation of a new AI regulator necessitated by Option 3 will also require significant investment in resources (including identifying those with the relevant capabilities) and time.

Given the above, at this early stage, our preference is for Option 2. The goal should be the development of a consistent and clear legal framework which avoids regulatory misalignment or conflicting regulatory settings.

As a broad observation, whilst the Consultation Paper draws heavily from the regulatory approach of the EU and Canada, which have taken a more prescriptive approach to AI, we would encourage further consideration of the approach taken by jurisdictions such as the UK⁶ and Singapore. This includes the provision of practical tools that incentivise the implementation of safe and responsible AI governance, rather than overly prescriptive regulation. For instance, Singapore has introduced the 'AI verify' tool to validate the performance of AI systems against AI ethics principles through standardised tests.

AI regulation must be coordinated and aligned with privacy, data and cyber security reforms

AI regulatory reforms need to be coordinated and aligned with concurrent reforms, particularly those in privacy, data governance and cyber security.

Privacy

As recognised by the Consultation Paper, in early September 2024 the Government introduced the *Privacy and Other Legislation Amendment Bill 2024 (Privacy Amendment Bill)*. This Bill introduces additional transparency requirements on organisations using personal information as part of automated decision-making (ADM) processes which could reasonably be expected to significantly affect rights or interests. The onus and costs on organisations to comply with this requirement is likely to be significant.

It is critical that privacy reform and the Guardrails are coordinated to ensure there is alignment, noting that such ADM requirements under the Privacy Amendment Bill are likely to feed into Mandatory Guardrails 5,6 and 7.

Data governance

Guardrails 2 and 3 require consideration of data governance. This intersects with existing legal obligations in relation to data collection, de-identification, use and retention. While reforms to these requirements have not been incorporated into the Privacy Amendment Bill, these areas suffer from significant complexity and have been earmarked for potential future reform by the Privacy Act Review.

Feedback from AICD members and industry experts is that there are significant challenges with interpreting, navigating and complying with Commonwealth, state and industry specific data laws. Gilbert + Tobin analysis indicates that there are approximately 100 laws, standards and enforceable guidelines applying to the retention and management of customer and business data in the financial services sector alone.⁷ These complexities will only be exacerbated with the use of AI, which relies on

⁶ The UK had taken a 'pro-innovation' approach to AI under the Conservative Government. However, the 17 July 2024 [King's Speech](#) flagged more interventionist approaches by the newly elected Labor Government, including introducing regulation covering "those working to develop the most powerful AI models."

⁷ Australian Financial Review, *Business navigates a maze of data obligations, law firm warns*, 20 March 2023.

vast amounts of data. Consideration should also extend to whether and how copyright law impacts the data available to train general-purpose models.

In addition to regulatory reform, we broadly support the Productivity Commission's suggestion of a comprehensive economy-wide national data strategy,⁸ noting that the [Data and Digital Government Strategy](#) applies only to the Australian Public Service (APS).

Cyber security

Advanced AI systems, particularly frontier AI systems, are more likely to present both significant opportunities and heightened risks for the cyber security resilience of Australian organisations. Any AI-specific regulation should recognise this dynamic and retain flexibility to respond to emerging cyber security risks.

Consideration should also be given as to whether specific obligations will need to be imposed on the development or deployment of AI which poses a serious risk to cyber security, national security or critical infrastructure, noting that President Biden's Executive Order on AI incorporated such requirements.⁹

6. Safe and responsible AI work plan

As recognised by the Consultation Paper, regulation alone is not sufficient to achieve the policy objective of maximising AI benefits while minimising harms. It is critical that regulation is accompanied by actions aimed at uplifting AI capability and governance skills and encouraging innovation.

We support the Government's 'Safe and Responsible AI work plan' and make the below additional observations/ comments.

Further focus on Pillar 3 encouraged

We welcome further details on the Government's plan for implementation of Pillar 3 (Supporting AI capability), which will be critical to maximising AI opportunities and enforcing regulation.

Research suggests that current levels of AI and digital competency of Australian business leaders and employees remain fairly low.¹⁰ However, the use of AI by and within organisations is estimated to exponentially increase - while 24% - 35% of Australian employees are currently using AI in their work,¹¹ this is estimated to grow to over 80% by 2030.¹² Notably, according to [IBM's 2023 Global AI Adoption Index](#), Australian organisations rank second lowest in active AI use.

⁸ Productivity Commission (January 2024) [Making the most of the AI opportunity: Research paper, no. 3 – AI raises the stakes for data policy](#) at page 17 (PDF page 19).

⁹ See Footnote 5 (President Biden's Executive Order on AI).

¹⁰ A [2023 KPMG and University of Queensland study](#) found that 59% of Australian respondents reported a low understanding of AI and when it's being used (page 6); 29- 36% of employees surveyed in a [2024 RMIT and Deloitte Access Economics study](#) stated that they did not have the relevant digital skills required, or that their skill was out of date (at page 4). Globally, a [2023 Boston Consulting Group study](#) found that 59% of executives stated that they have limited or no confidence in their executive team's proficiency in the use of Generative AI.

¹¹ Estimates of current AI use by employees vary. A [2023 global KPMG and University of Queensland study](#) found 24% of Australian respondents stated that AI is used in their employing organisation (Figure 29 at page 46 or PDF page 48), whilst a [PersonKelly study](#) found that 35% of Australian employees used AI (at page 4 or PDF page 3). Another study by Deloitte found that 32% of employees are using Generative AI specifically, with nearly two-thirds believing their managers are unaware of their use (known as 'shadow AI')(cited in the [2023 Australian Computer Society and Deloitte study \(ACS and Deloitte Study\)](#) at page 12).

¹² A [2024 Amazon Web Services survey of Australian workers \(AWS Report\)](#) found that 86% of employees expect to use AI in their daily work by 2028, of which 25% expect to use it "extensively"(at page 5). The [ACS and Deloitte Study](#) concluded that 86% of occupations have skills that will be affected by AI technologies, and 25% of all work time will be affected, and 52% of occupations will have at least 20% of their work time impacted by AI (at page 21).

Notwithstanding the growing demand for AI and digital skills across the Australian economy,¹³ companies have struggled to provide adequate AI training to their employees.¹⁴ The inability of existing employee upskilling and new employee entrants to keep pace with rising digital and AI workforce demand has led experts to predict a 'digital worker shortfall' of anywhere between 370,000¹⁵ and 495,687.¹⁶

[Jobs and Skills Australia](#) has recognised the need for education policy to address the growing importance of digital and AI skills. It has called out digital transformation and the emergence of AI as a "key megatrend" requiring a response.¹⁷ We appreciate that there are many existing initiatives to uplift digital skills, such as the [Future Skills Organisation](#) (within Jobs and Skills Australia) and the [Australian Digital Capability Framework](#). Consideration should be given as to how AI skills can be incorporated into these initiatives, as well as to the role that domestic investment and capability building in AI plays in the [Future made in Australia](#) strategy.

We would encourage the Government's 'Pillar 3' to also address strategies to: (1) uplift the skills of existing workforce participants; (2) develop the skills of new workforce entrants; and (3) attract overseas AI talent to fill any shortfall and/or to assist in the uplift. One approach to address these issues may be to design and implement a National Digital Skills Strategy.

Uplifting AI capability amongst government and regulators should also be prioritised to ensure existing laws and any AI-specific laws are appropriately designed and adequately enforced.

AICD role in supporting board-level AI understanding and capability

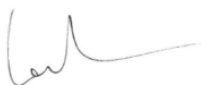
The AICD sees itself as playing a key role in supporting and promoting best-practice governance and uplifting AI capability (i.e. Pillars 2 and 3) at the director and board level.

In June 2024, the AICD, in collaboration with the Human Technology Institute (HTI) released an [AI governance for directors' resource suite](#), which has had over 12,500 downloads to date. This is supported by the [AICD's Governance of AI webinar series](#) and [content](#), including featuring governance of AI in member communications and major AICD events. To uplift director understanding the AICD is also piloting [an AI fluency for directors course](#) in collaboration with the University of Sydney. We would be pleased to discuss these initiatives with the Department and explore potential areas for collaboration to support effective AI governance.

7. Next Steps

If you would like to discuss these matters further, please contact Christian Gergis, Head of Policy at cgergis@aicd.com.au or Anna Gudkov, Senior Policy Adviser at agudkov@aicd.com.au.

Yours sincerely,



Louise Petschler GAICD

General Manager,
Education & Policy Leadership

¹³ According to the [AWS Report](#), hiring AI-skilled talent is a priority for 63% of Australian employers, of which 75% cannot find the AI talent they need (at page 5). According to a [NAIC 2023 AI ecosystem report](#), Australia is among the global leaders in terms of AI job postings, with 1.2% of all job postings in 2022 being AI-related. Demand for AI jobs has also been going faster in Australia relative to overseas, with the share of AI-related job postings in Australia increasing by more than 7 times between 2014 and 2022 (at page 24, PDF page 26).

¹⁴ According to the [AWS Report](#), 73% of employers state they do not know how to implement an AI workforce training program, while 76% of workers say they are not sure of what AI training programs are available to them (page 5).

¹⁵ See the Digital Skills Organisation (June 2023) [Growing Australia's digital workforce Report](#) at page 8.

¹⁶ See the [ACS and Deloitte Study](#) at page 28.

¹⁷ Jobs and Skills Australia (2023), [Annual Jobs and skills report 2023](#) at page 12.