Deloitte.

# Safe and responsible AI in Australia

**Deloitte response to Department of Industry, Science and Resources**

July 2023

Deloitte.

Deloitte Touche Tohmatsu
ABN 74 490 121 060
477 Collins Street
Melbourne, VIC 3000
Australia
P: +61 3 9671 7000
www.deloitte.com.au

The Honourable Ed Husic MP, Minister for Industry and Science
Department of Industry, Science and Resources
10 Binara Street
Canberra, ACT 2601
Australia

26 July 2023

Dear Minister Husic

**Re: Safe and Responsible AI in Australia**

Deloitte recognises the enormous potential that Artificial Intelligence (AI) presents and is committed to being part of safely harnessing this opportunity for Australia, allowing innovation with confidence.

We base our submission on findings from our global State of AI report[1], consultation across our global network of firms, cross sector experience working with clients developing AI solutions, and our deep subject matter expertise in technology, regulation, risk management and governance. We have reviewed the emerging global regulation, principles and best practices, and advocate for a proportionate and pro-innovation approach to safe and responsible AI in Australia.

The questions posed in the discussion paper can be distilled to two foundational issues – what regulatory framework should be applied to AI and what AI systems fall within this regulatory framework. Answers to these questions will provide the 'north star' by which Australia should govern AI to harness its enormous potential, while protecting its citizens, values and unique environment.

It is our view that Australia should:

**Adopt a risk-based framework for AI regulation that supports the nuanced and proportionate management of risk and complements existing regulation**

Deloitte supports a risk-based regulatory framework that is focussed on risk posed by AI to individuals, society and the environment rather than targeting specific technologies. This enables a balanced approach to weighing up the benefits versus the potential risks. This should be underpinned by a set of principles and rules to drive consistency across sectors and organisations. These principles will also provide the flexibility to harmonise with existing domestic legislation and compatibility with emerging international regulation.

**Develop a definition of AI that supports the objectives of the overarching regulatory framework and is flexible to encompass future AI technologies**

AI should be defined in context of the regulatory framework by which it will be governed and the definition should be adaptable to future developments of the technologies. AI should be defined generally to encompass all socio-technical systems that have the potential to cause harm, aligned to the risk-based approach to regulation that is recommended here.

Deloitte proposes the following definition which meets these objectives:

*Artificial Intelligence refers to a computer-based system that can, for a given set of objectives, generate outputs such as content, predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.*

**Develop an AI code of practice to support safe and responsible AI**

Resolution of the foundational issues, through design and operationalisation of nuanced and effective AI regulation, will take time. Deloitte recommends that, in the meantime, Australia develops an AI code of practice to provide practical guidance on how to achieve safe and responsible AI.

We thank the Department of Industry, Science and Resources for the opportunity to contribute to this critically important national policy discussion.

Yours sincerely

Dr. Elea Wurth
Trustworthy AI lead

Dr Kellie Nuttall
AI Institute Lead

[1] Please refer to Deloitte (2022d).

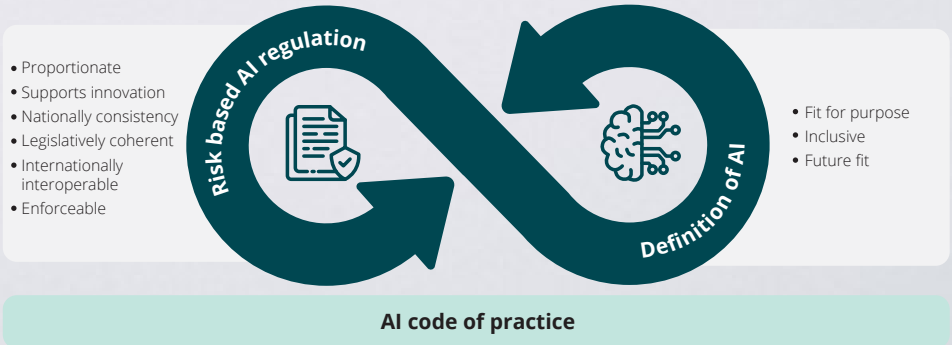
# Safe and responsible AI

Artificial intelligence (AI) is rapidly emerging as the defining technology of our time. It has transformed industries ranging from ride-share to radiology, with further developments expected to deliver economic and societal benefits across the spectrum of industries and social domains. By enhancing predictive capabilities, optimising operations and resource allocation, and personalising service and citizen interactions, AI will contribute to positive social and environmental outcomes on a significant scale while offering improved delivery of government services, and the potential to provide significant competitive advantage to business and the Australian economy.

However, there are increasing concerns about the potential negative impacts of AI. These range from privacy concerns and protection of intellectual property, to threats to national security and critical infrastructure, and undermining the human experience of fulfilling work and social connection. Given the pace of development, defining the role of regulation and its application to support safe and responsible AI has become increasingly urgent across the globe.

AI technologies range in complexity and can be applied in many different ways. This results in a wide variation in the risk posed, from the insignificant to the severe[2]. For example, using a generative AI chatbot to produce a summary of a long form magazine article presents very different risks to using the same technology to provide medical advice. This is also the case for more simple technologies, such as the significant difference in the risk posed by a set of coded logic statements applied to calculate a family budget as compared to the use of coded logic to deliver the Robodebt scheme[3].

*Figure 1. Meeting the needs of safe and responsible AI in Australia*
*Source: Deloitte Australia*




[2] Organisation for Economic Co-operation and Development (2022).
[3] The Royal Commission into the Robodebt Scheme (2023).

# Regulation of AI

## Deloitte supports flexible and proportionate regulation of AI

There is currently no general regulation of AI in Australia, however, that is not to say that AI is unregulated. Certain risks posed by AI are governed by both general (e.g., privacy, copyright and discrimination law) and sector-specific regulations (e.g., therapeutic goods and financial services). However, there are gaps between the risks that AI poses and existing regulatory frameworks, which is ever widening with the rapid advancement of the technology.

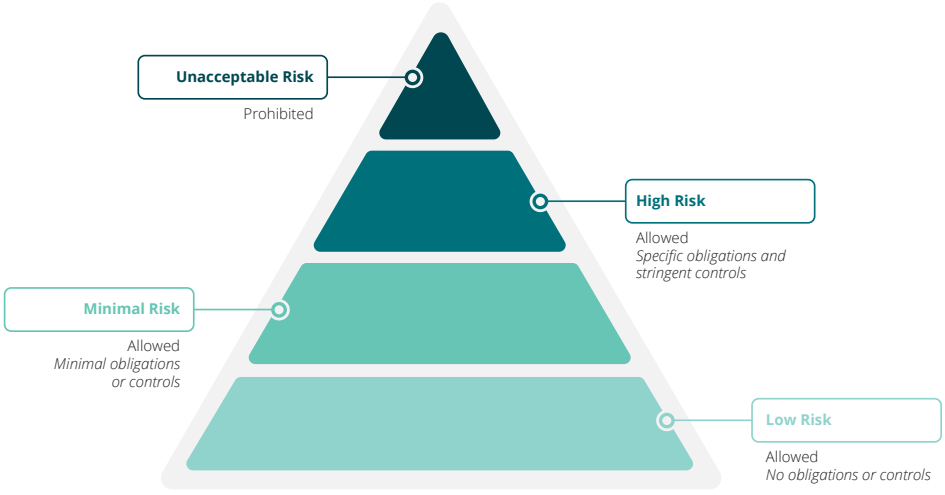
Regulation should strike a balance between fostering innovation and incentivising the responsible use of AI. Safe and responsible AI will be achieved most efficiently and effectively through regulation that is *flexible enough to cover the wide array of existing and emerging AI technologies* and proportionate to the contextualised risk that these technologies pose. Further, principles of AI regulation should be both consistent across sectors and harmonised with existing laws and sector-specific legislation such as data protection and privacy, online safety, anti-discrimination and Australian consumer law.

### Risk-based regulation

Deloitte supports risk-based AI regulation which is a regulatory framework designed to govern the development, deployment, and use of AI systems based on their *potential risks* to individuals, society, and the environment.

The key principle of risk-based AI regulation is the alignment of regulatory measures and compliance obligations with the level of risk posed by AI systems. It recognises that not all applications of AI carry the same level of risk, and therefore, regulatory efforts and compliance burdens should be proportionate and targeted. This approach allows prioritisation of risk management and compliance resources by both AI actors and regulators on higher risk applications while minimising undue burdens on low-risk AI systems (see Figure 2). Risk-based AI regulation aims to strike a balance between fostering innovation and ensuring the responsible and ethical use of AI systems.

*Figure 2. A risk-based framework for AI regulation*
*Source: Deloitte Australia*



## Risk-based AI regulation typically include the following components:

01

### Risk Assessment

The regulatory framework defines the risk assessment criteria associated with AI systems by considering factors such as the system's purpose, capabilities, potential impact on individuals and society, and the context of its use. This assessment helps identify the level and nature of risks involved.

02

### Risk Classification

Based on the risk assessment, AI systems are categorised into different risk classes or levels, ranging from low to high risk. The classification may consider factors like potential harm, vulnerability of the affected population and the likelihood of adverse outcomes.

03

### Regulatory Requirement

Specific requirements and guidelines are developed for each risk class, tailoring them to address the identified risks. These requirements may cover areas such as governance, controls, transparency, explainability, fairness, accountability, safety and security.

04

### Compliance and Oversight

Organisations and individuals developing or deploying AI systems are required to comply with the applicable regulatory requirements based on the risk class of their systems. Regulation may include audits, inspections, or other oversight activities to ensure compliance, detect non-compliance and take appropriate enforcement actions if needed.

05

### Continuous Monitoring and Adaptation

Implement continuous monitoring and evaluation of AI systems and their impact. As technology evolves and new risks emerge, regulatory frameworks need to be adaptable and updated to effectively address those risks.

The higher the risk, the stricter the rule. Risk-based AI regulation recognises that rules and compliance procedures for AI applications that do not present genuine risks can result in higher costs and burdens, without providing real benefits, therefore regulatory obligation should be proportional to risk. AI developers and users should be, for example, free to experiment with new technologies in sandbox environments and release low-risk applications. Innovation in high-risk AI systems, such as those deployed in critical infrastructure or that impact employment and education opportunities, can also be supported by safeguards and appropriate levels of governance and control.

### Horizontal application

A risk-based regulatory framework should be implemented consistently and coherently across sectors, organisation size and AI maturity to govern the development, deployment and use of AI systems. This 'horizontal' application of AI regulation can help establish a cohesive and nationally consistent regulatory framework for AI systems, across the public and private sectors. By addressing common challenges, setting universal standards and promoting collaboration, it facilitates a level playing field for the regulatory expectations and obligations of AI actors.

A consistent and coherent approach can also enhance consumer trust and confidence. Consumers are more easily able to understand their rights and develop confidence in their own ability to balance risk and benefit offered by AI systems through consistent standards and safeguards. This trust is essential for the widespread adoption of AI technologies and the realisation of their benefits.

Horizontal application of AI regulation does not replace existing legislation. Rather this approach seeks to harmonise with existing Australian general regulation (e.g., data protection and privacy, consumer law, online safety, copyright law etc) and sector-specific regulation (e.g., therapeutic goods and financial services). The intent is to close the gap between existing Australian regulation and the emerging risk that AI systems pose.

### Global compatibility

Horizontal application of risk-based regulation is consistent with the European Union (EA) AI Act which is the forerunner in global AI regulation and emerging as the dominant approach among comparable jurisdictions. Australian implementation of a compatible regulatory framework could help facilitate cross border collaboration on AI development through shared standards and global management of challenges associated with AI. Jurisdictional compatibility also reduces regulatory complexities for organisations operating in multiple countries, enabling smoother cross-border AI deployments and reducing the risk of actors gaming jurisdictional regulation.

### Regulatory oversight

An oversight body should be established to implement and administer a regulatory framework for safe and responsible AI, ensuring consistency across federal and state levels. The specific structure, powers and processes of an AI oversight body would depend on the legal and regulatory framework to be implemented.

Depending on the framework, the role of an AI oversight body might include developing policies, regulations and guidelines that govern the development and use of AI technologies, monitoring and application of sanctions to enforce compliance with the established regulations and guidelines, development of certification or accreditation programs to verify compliance with established standards, and providing an avenue for consumer complaints and users to seek review of a decision made by AI systems.

Given the global nature of AI, the oversight body should actively engage in international collaboration opportunities. It should work with international regulators, share best practices and contribute to the development of international standards and norms to ensure consistency and interoperability across borders.

The body's effectiveness would rely on its ability to strike a balance between supporting innovation and protecting the rights, safety and welfare of individuals and society and environment.

# Definition of AI

The Australian definition of AI must be developed in context of the regulatory approach by which it is to be governed. That is, if it were to be subject to narrow, rules-based regulation of specific technologies, AI should be defined precisely and explicitly. If, as recommended in this submission, it is to be regulated according to the level of risk posed to individuals, society and the environment, the definition should be general and encompas all AI systems that have the potential for negative outcomes.

The definition of AI proposed in the discussion paper is largely aligned to definitions proposed by the EU AI Act[4] and OECD AI Principles[5], with the critical distinction of stipulating that output must be generated without explicit programming:

*"Artificial Intelligence" refers to an engineered system that generates predictive outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives or parameters without explicit programming. AI systems are designed to operate with varying levels of automation.*

Unlike our international counterparts, this definition restricts AI to machine learning – and therefore with this definition, Australian AI regulation would only apply to machine learning systems and not the broader category of automated decision systems that have the potential to pose risk.

Deloitte recommends that the Australian definition of AI:

01

Replaces 'engineered' with 'computer-based' to narrow the breadth of potential system inclusion to information processing systems only.

02

Excludes 'without explicit programming' to expand the definition to include automated systems based on coded logic statements. This inclusive approach is aligned to the dominant emerging definition of AI across comparable jurisdictions and international policy organisations[6].

03

Excludes 'human-defined' to be flexible to future post-generative AI systems. Generative AI has brought us closer to general AI which has the potential to develop parameters and to create and pursue objective options that are not explicitly defined by humans.

[4] Please refer to European Commission (2021).
[5] Please refer to Organisation for Economic Co-operation and Development (2019).
[6] Please refer to European Commission (2021), Organisation for Economic Co-operation and Development (2019), Information Commissioner's Office (2022).

"Artificial Intelligence" refers to a computer-based system that can, for a given set of objectives, generate outputs such as content, predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.

# AI code of practice

The design and operationalisation of nuanced and effective AI regulation will take time. In the interim, Deloitte recommends that Australia develops an AI code of practice outlining a set of guidelines, principles, and standards that can be used to govern the AI system lifecycle.

An AI code of practice should start from the foundational position that AI is a socio-technical system that must be lawful, ethical and technically robust. This approach emphasises that it is the dynamic human-machine interaction that creates the potential for risk and this risk should be managed across people, process and technology. A key finding of the Deloitte State of AI in Enterprises report[7] was to ensure ethical and quality application of AI – the entire operating model may need to change to accommodate the unique capabilities of intelligent machines. Workflows and roles should be re-evaluated to manage risk and achieve new value.

Design of an AI code of practice should draw from existing approaches to AI risk management and governance and align to the emerging global principles on responsible AI, for example those outlined in Australia's AI Ethics principles[8]:

**Human, societal and environmental wellbeing**

AI systems should benefit individuals, society and the environment.

**Human-centred values**

AI systems should respect human rights, diversity, and the autonomy of individuals.

**Fairness**

AI systems should be inclusive and accessible and should not involve or result in unfair discrimination against individuals, communities or groups.

**Privacy protection and security**

AI systems should respect and uphold privacy rights and data protection, and ensure the security of data.

**Reliability and safety**

AI systems should reliably operate in accordance with their intended purpose.

**Transparency and explainability**

There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI and, can find out when an AI system is engaging with them.

**Contestability**

When an AI system significantly impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or outcomes of the AI system.

**Accountability**

People responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems and human oversight of AI systems should be enabled.

Features of an AI code of practice should support safe and responsible AI across the design, development, deployment and monitoring phases of AI. The AI code of practice should be flexible enough to evolve with future technologies and integrate with different possible regulatory frameworks.

[7] Please refer to Deloitte (2022d).
[8] Please refer to Department of Industry, Science and Resources(2019).

# Resources

Deloitte. 2022a. Trustworthy Artificial Intelligence. Refer to: https://www.deloitte.com/au/en/Industries/consumer/analysis/trustworthy-artificial-intelligence. Last accessed: 21/07/2023

Deloitte. 2022b. The Implications of Generative AI for Businesses. Refer to: https://www2.deloitte.com/us/en/pages/consulting/articles/generative-artificial-intelligence.html. Last accessed: 21/07/2023

Deloitte. 2022c. The AI Dossier. Refer to: https://www.deloitte.com/au/en/services/consulting/perspectives/artificial-intelligence-dossier.html. Last accessed 21/07/2023

Deloitte. 2022d. State of AI in Enterprises 2022. Refer to: https://www2.deloitte.com/us/en/pages/consulting/articles/state-of-ai-2022.html. Last accessed: 21/07/2023

Deloitte. 2023. 2023 – 2030 Australian Cyber Security Strategy. Refer to: https://www.deloitte.com/au/en/services/risk-advisory/perspectives/2023-2030-australian-cyber-security-strategy-deloitte-austra.html. Last accessed: 21/07/2023

Organisation for Economic Co-operation and Development. 2019. OECD AI Principles. Refer to: https://oecd.ai/en/ai-principles. Last accessed: 21/07/2023

Organisation for Economic Co-operation and Development. 2022. OECD Framework for the Classification of AI systems. Refer to: https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9eca-en.htm. Last accessed: 21/07/2023

European Commission. 2021. Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. Refer to: https://artificialintelligenceact.eu/the-act/. Last accessed: 21/07/2023

The Royal Commission into the Robodebt Scheme. 2023. Report. Refer to: https://robodebt.royalcommission.gov.au/system/files/2023-07/report_of-the-royal-commission-into-the-robodebt-scheme.pdf Last accessed: 21/07/2023

Department for Science, Innovation and Technology. 2023. A pro-innovation approach to AI regulation. Refer to: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach Last accessed: 21/07/2023

National Institute of Standards and Technology. 2022. AI Risk Management Framework. Refer to: https://www.nist.gov/itl/ai-risk-management-framework. Last accessed: 21/07/2023

Information Commissioner's Office. 2022. Explaining Decisions made with Artificial Intelligence – Definitions. Refer to: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence/part-1-the-basics-of-explaining-ai/definitions/. Last accessed: 21/07/2023

Department of Industry, Science and Resources. 2019. Australia's AI Ethics Principles. Refer to: https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles. Last accessed: 26/07/2023

# Contacts

**Dr Elea Wurth**
Partner Risk Advisory
Tel: +61 3 8486 1732
ewurth@deloitte.com.au
–
**Dr Kellie Nuttall**
Partner Consulting
Tel: +61 7 3308 7075
knuttall@deloitte.com.au

Deloitte.