Australia's Approach to Responsible AI in a Global Context

THADDEUS MARTIN

Australia's Approach to Responsible AI in a Global Context

Thaddeus Martin Consulting

January 2025

Copyright Statement

'Australia's Approach to Responsible AI in a Global Context' is a publication of: Thaddeus Martin Consulting, 74-76 Campbell Street, SURRY HILLS NSW 2010, Australia.

This work is protected under the **Copyright Act 1968 (Cth)** and is owned by **Thaddeus Martin Consulting**.

You are free to **use, reproduce, and distribute this work for non-commercial purposes**, provided proper attribution is given to the original authors and Thaddeus Martin Consulting. Attribution should include the title of the work, the authors' names, the publisher Thaddeus Martin Consulting, and the year of publication.

For any **commercial use or adaptations**, prior written permission is required.

inquiries or permissions contact **Darren Conlon** at
darrenconlon@thaddeusmartin.com

How to cite this publication

Huang, A., Conlon, D., &, Tayco, D. (2025). *Australia's approach to responsible AI in a global context*. Thaddeus Martin Consulting.

https://www.thaddeusmartin.com/insights/

Contents

1. Definition and Significance of Responsible AI

Definition

"Responsible AI (RAI) refers to the design, development, deployment and governance of AI in a way that respects and protects all human rights and upholds the principles of AI ethics through every stage of the AI lifecycle and value chain."[1]

Significance

- The rapid adoption of AI in the workplace has highlighted an urgent need to ensure its development, use, and governance are conducted responsibly.[2]
- There are gaps in RAI coverage including without limitation in protections for:
  - Vulnerable or marginalised groups.
  - Workers whose roles are negatively affected or undermined by AI.
  - Inclusion and equality (*e.g.*, gender equality, cultural and linguistic diversity).
  - The safety, security, and reliability of AI systems.[3]
- Additionally, the UNPRI recommend including RAI considerations in investment decision-making.[4]

Outstanding issues

- The promotion, use, and advancement, of AI is generally underpinned by unenforceable principles.
- There is a paucity of consistent concrete approaches to promulgating RAI.
- Countries are at different stages in the development and implementation of AI policies and regulations.
- To date most countries and international organisations (the EU, OECD, UNESCO, *etc.*) have adopted a voluntary or recommendations led approach.
- In Australia, the Federal Government has issued comprehensive guidelines for ethical and secure AI use but has not yet promulgated legislation on it.

---

[1] Adams, R., Adeleke, F., Florido, A., de Magalhães Santos, L. G., Grossman, N., Junck, L., & Stone, K. (2024). Global Index on Responsible AI 2024 (p.9). Global Center on AI Governance. https://www.global-index.ai

[2] Adams, R., Adeleke, F., Florido, A., de Magalhães Santos, L. G., Grossman, N., Junck, L., & Stone, K. (2024). Global Index on Responsible AI 2024. Global Center on AI Governance. https://www.global-index.ai

[3] *Ibid.*

[4] UNPRI. (2022). Collaborative Engagement to Improve Ethical and Responsible AI. https://collaborate.unpri.org/group/11711/about

2. Australia's RAI performance

Australia's performance rated amongst current RAI scoring and ranking models.

(a) Global Index on Responsible AI (GIRAI)

GIRAI is a comprehensive tool that measures progress toward responsible AI across 138 countries. It evaluates various dimensions, including human rights, governance, and AI capabilities, providing a global benchmark for responsible AI practices.[5]

Criteria evaluated		
Criterion 1: **Human Rights and AI**	**Criterion 2:** **Responsible AI Capacities**	**Criterion 3:** **Responsible AI Governance**
• Gender equality. • Data protection and privacy. • Public participation and awareness. • Bias and unfair discrimination. • Protection of children's rights. • Labor protection and the right to work. • Cultural and linguistic diversity.	• Competition authorities' roles in overseeing AI. • Public sector skills development. • International cooperation on AI standards and frameworks.	• National AI policies. • Use of impact assessments to measure AI's effects. • Human oversight and determination in decision-making processes. • Accountability mechanisms for responsible deployment. • Proportionality and avoidance of harm in AI implementation. • Transparency and explainability in AI systems. • Safety, accuracy, and reliability of AI systems. • Access to remedies and redress for AI-related grievances.

Score: 56.22[6]

Ranking: 10th globally

Strengths:

    (a) Performs well in governance frameworks, policies, and accountability measures.

    (b) Demonstrates leadership in gender equality, privacy, and data protection within AI systems.

Weaknesses:

    (a) Lags behind leading nations like the Netherlands (86.16) and Germany (82.77) in technical capacities and overall responsible AI maturity.

---

[5] Adams, R., Adeleke, F., Florido, A., de Magalhães Santos, L. G., Grossman, N., Junck, L., & Stone, K. (2024). Global Index on Responsible AI 2024. Global Center on AI Governance. https://www.global-index.ai

[6] The 3 pillars of the RAI ecosystem *Government Frameworks*, *Government Actions*, and *Non-State Actors* are benchmarked against the Global Index Evaluation Criteria to generate the 'Index Score'.

(b) Government AI Readiness Index (Oxford Insights)

This index evaluates governments' preparedness to implement AI in public services, considering factors like governance, infrastructure, and innovation capacity.[7]

Criteria evaluated		
Criterion 1: **Governance**	**Criterion 2:** **Infrastructure**	**Criterion 3:** **Innovation**
• Development and implementation of national AI strategies. • Ethical AI policies and public leadership.	• Cloud computing availability. • Broadband penetration and digital access.	• Research output and publications. • Industry engagement and partnerships. • Development of AI ecosystems.

Score: Not explicitly provided for individual metrics.

Ranking: 8th globally

Strengths:

(a) Australia's advanced AI governance and ethical frameworks boosted its ranking.

(b) Significant investments in research and innovation.

Weaknesses:

(b) Insufficient digital skills development within the public sector.

---

[7] Oxford Insights. (2024). Government AI Readiness Index 2024. https://oxfordinsights.com/ai-readiness/ai-readiness-index/

(c) Stanford Global AI Power Index

Developed by Stanford University's Institute for Human-Centered AI, this tool assesses the AI capabilities of 36 nations using 42 indicators, including research output, investment, and patents. The U.S. leads in virtually every pillar, including machine learning releases, private AI investment, and responsible AI research.[8, 9]

Criteria evaluated		
Criterion 1: **Research Output**	**Criterion 2:** **Investment**	**Criterion 3:** **AI Capabilities**
• Volume and quality of AI-related publications across academic and industry domains. • Areas of focus include machine learning, robotics, and natural language processing (NLP).	• Evaluation of public and private sector funding in AI development. • Commercialisation efforts for AI innovations.	• Performance in NLP, robotics, and other AI benchmarks. • Development of scalable AI models for public and private sector use.

Score: No explicit score available for Australia.

Ranking: Not explicitly ranked for Australia.

Strengths:

(a) Australia is recognised for significant contributions in NLP and machine learning.

(b) High-impact research publications and active academic-industry collaborations.

Weaknesses:

(a) Limited visibility in global rankings compared to leading countries like the United States.

---

[8] Stanford Institute for Human-Centered Artificial Intelligence. (2022). Annual report 2022. https://hai.stanford.edu
[9] Stanford Institute for Human-Centered Artificial Intelligence. (2023). Annual report 2023. https://hai.stanford.edu

(d) Australian Responsible AI Index

This index focuses specifically on Australian organisations, categorising them into four maturity levels: Emerging, Developing, Implementing, and Leading.[10]

Criteria evaluated		
Criterion 1: **Responsible AI Maturity**	**Criterion 2:** **Policy Adherence**	**Criterion 3:** **Ethical AI Practices**
• Categorisation of organisations into Emerging, Developing, Implementing, and Leading groups.	• Evaluation of compliance with governance and ethical AI principles.	• Transparency, fairness, accountability, and inclusivity in AI systems.

Score: Mean score of 44/100 for Australian organisations.

Ranking: No global ranking (national focus).

Strengths:

(a) Growing awareness of responsible AI practices among Australian organisations.

(b) Evidence of improvement in governance and compliance mechanisms.

Weaknesses:

(a) Most organisations remain in the Emerging or Developing categories, with low maturity in ethical AI implementation.

---

[10] Fifth Quadrant. (2024). Australian Responsible AI Index 2024. https://www.fifthquadrant.com.au/responsible-ai-in-australia-2024-index

3. Global approaches to RAI

a. OECD AI Key Principles

The aims of the OECD Values-based Key Principles[11] underpinning the use of RAI are:

- Promoting **inclusive growth, sustainable development, and well-being**.
- Upholding **human rights and democratic values**, including fairness and privacy.
- Ensuring **transparency and explainability** of AI systems.
- Guaranteeing **robustness, security, and safety** in AI operations.
- Emphasising **accountability** across all stages of AI system development and deployment.

Recommendations for Policy Makers:

- Invest in AI research and development.
- Foster an inclusive AI-enabling ecosystem.
- Fashion an enabling, interoperable, governance and policy environment for AI.
- Build human capacity and prepare for labour market transition.
- Engage in international co-operation for trustworthy AI.

Globalisation

OECD AI Key Principles have been adopted by many countries and organisations to create their AI policies and frameworks.[12]

---

[11] OECD. (2024). *OECD AI Principles Overview* (2nd ed.). https://oecd.ai/en/ai-principles

[12] Countries adhering to OECD Principles are Australia, Austria, Belgium, Canada, Chile, Colombia, Costa Rica, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Türkiye, United Kingdom, US.

b. Comparative Exemplars

Asia Pacific	Europe	North America
South Korea (voluntary)[13] • First country to develop nation-level AI legislation. • The *Framework Act on Intelligent Informatisation* (2020), Article 56, allows state/local governments to survey and assess AI's social impact. • Primarily applies to the public sector.	**EU (regulatory)** • The *General Data Protection Regulation (GDPR)*[14] mandates ex-ante impact assessments (Art. 35) for high-risk AI applications (e.g., biometric data processing, automated decision-making with legal effects). • The Digital Services Act (DSA)[15] governs online platforms, including AI-driven services.	**US (non-binding)[16]** • The Blueprint for an AI Bill of Rights provides guidance but remains non-binding.[17] • A Trustworthy AI Executive Order requires federal agencies to follow nine AI principles. • Certain state laws impose transparency and validity requirements, particularly in areas like pretrial risk assessments.
Singapore (voluntary)[18] • The *Personal Data Protection Commission (PDPC)* Model AI Governance Framework provides AI governance guidelines emphasising fairness, transparency, and accountability.	**UK (principles based)[19]** • Adopts AI governance principles rather than rigid laws, focusing on ethical AI development.	**Canada (proposed regulatory)[20]** • The *Artificial Intelligence and Data Act (AIDA)* is a proposed law to regulate high-impact AI at the federal level. • Currently, AI governance relies on a voluntary code for businesses.

[13] Adams, R., Adeleke, F., Florido, A., de Magalhães Santos, L. G., Grossman, N., Junck, L., & Stone, K. (2024). Global Index on Responsible AI 2024. Global Center on AI Governance. https://www.global-index.ai

[14] European Parliament, & Council of the European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation).* Official Journal of the European Union, L 119, 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj

[15] European Parliament, & Council of the European Union. (2022). *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act).* Official Journal of the European Union, L 277, 1–102. https://eur-lex.europa.eu/eli/reg/2022/2065/oj

[16] The White House. (2022). *AI Bill of Rights.* https://www.whitehouse.gov/ostp/ai-bill-of-rights/safe-and-effective-systems-3/

[17] Redacted for review by Executive Order of the President of the United States 23 January 2025.

[18] Personal Data Protection Commission. (2020). *Model artificial intelligence governance framework (2nd ed.).* Infocomm Media Development Authority. https://www.pdpc.gov.sg/Help-and-Resources/2020/01/Model-AI-Governance-Framework

[19] UK Government. (2023). *A pro-innovation approach to AI regulation: Government response.* Department for Science, Innovation & Technology. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach

[20] Government of Canada. (2022). *Artificial Intelligence and Data Act (AIDA) - Part of Bill C-27, Digital Charter Implementation Act, 2022.* https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading

c. Which *model should Australia aspire to?*

Australia can enhance its global standing by adopting proactive, enforceable standards aligned with the OECD and EU whilst tailoring regulations to local priorities, such as Australian and Indigenous data sovereignty.

Australia

- o Currently, Regulation of AI is *voluntary, so* Australia does not have any specific statutes or regulations directly regulating AI.
- o According to:
  - The 2024 Global Index on Responsible AI[21] Australia is ranked 10th globally in sustaining a contemporary legal and ethical ecosystem conducive to incorporating RAI, with a score of 56.22 putting it amongst the second-best group. "Countries that performed well were able to ***demonstrate a wide range of governance mechanisms – including sector specific policies and legislative frameworks*** – to safeguard human rights and advance RAI development and use."[22]
  - The Government AI Readiness Index (Oxford Insights)[23] Australia is ranked 8th globally in governments' preparedness to implement AI in public services, considering factors like governance, infrastructure, and innovation capacity. *Countries that performed well **demonstrated strong investments in digital infrastructure, innovative public services, and comprehensive governance frameworks** to integrate AI responsibly across various sectors*.
  - The Stanford Global AI Power Index,[24,25] Australia is not specifically ranked but s recognised for its contributions to research and innovation, particularly in natural language processing (NLP). *Countries that performed well **excelled in AI-related research output, significant advancements in natural language processing, and fostering collaboration between academia and industry***.

---

[21] Adams, R., Adeleke, F., Florido, A., de Magalhães Santos, L. G., Grossman, N., Junck, L., & Stone, K. (2024). Global Index on Responsible AI 2024. South Africa: Global Center on AI Governance. https://girai-report-2024-corrected-edition.tiiny.site/

[22] *Ibid* p.29.

[23] Oxford Insights. (2024). Government AI Readiness Index 2024. https://oxfordinsights.com/ai-readiness/ai-readiness-index/

[24] Stanford Institute for Human-Centered Artificial Intelligence. (2022). Annual report 2022. https://hai.stanford.edu

[25] Stanford Institute for Human-Centered Artificial Intelligence. (2023). Annual report 2023. https://hai.stanford.edu

- The Australian Responsible AI Index 2024 Australian Organisations scored out of 100 when categorised into four maturity levels: Emerging (0-24), Developing (25-49), Implementing (50-69), and Leading (70+), indicating significant room for improvement.[26] *Organisations that performed well showcased **advanced adherence to ethical AI practices, robust policy frameworks, and a commitment to achieving higher levels of responsible AI maturity***.

o Clearly, there is room for improvement. To bridge the gap between principles and enforceable frameworks, a phased regulatory approach could be adopted, beginning with mandatory requirements for high-risk AI applications, transitioning to sector-specific legislation, and culminating in comprehensive national AI laws, that include without limitation:

- AI Ethics Key Principles[27] introduced in 2019 (adhering to OECD Key Principles).
- The 10 Guardrails[28] published in 2024 – proposed mandatory obligations for high-risk AI settings.

---

[26] Fifth Quadrant. (2024). *Australian Responsible AI Index 2024*. https://www.fifthquadrant.com.au/responsible-ai-in-australia-2024-index

[27] Australian Government, Department of Industry, Science and Resources. (2019). *Australia's AI Ethics Principles*. https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles

[28] Australian Government, Department of Industry, Science and Resources. (2024). *The 10 Guardrails*. https://www.industry.gov.au/publications/voluntary-ai-safety-standard/10-guardrails

4. Concerns underpinning AI

The importance of RAI is exemplified by three broad concerns demonstrating the need for Australian policy development in this area.[29, 30, 31, 32]

Category	Definition & Significance
Extraction of Data	**Definition:** Extraction of data refers to collecting information from different sources for AI development. **Significance:** Risks of infringement of Intellectual Property, and privacy and security concerns. For example, training datasets extracted from social media platforms could inadvertently capture sensitive personal information without consent.
Extraction of Minerals	**Definition:** The process or procedure of withdrawing or retrieving minerals from the earth for human use. **Significance:** The extraction of minerals and resources to create and power AI systems and support networks has potential and actual environmental impacts, which can also undermine Environmental, Social, and Governance (ESG) policies.
Extraction of Labour (*focusing on "exploitation" of labour during AI development*)	**Definition:** Fair labour practices should include minimum wage standards, regular audits of working conditions, contractual guarantees for safe and ethical work environments, *etc*. Exploitation refers to an unfair working environment where the employer benefits from unethical or illegal treatment towards their employees. **Significance:** Engagement of human labour in the process of developing and training AI systems carries a credible risk of exploitation. For example, data-labelling is necessary in the development of AI models, which requires significant human labour. Data-labellers are found to be working long-hours and are significantly under-paid.

[29] UNESCO. (2021). *Recommendation on the ethics of artificial intelligence.* United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000379920

[30] OECD. (2019). *OECD principles on artificial intelligence.* Organisation for Economic Co-operation and Development. https://oecd.ai/en/ai-principles

[31] Maslej, N., Fattorini, L., Perrault, R., Parli, V., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., & Clark, J. (2024). *The AI Index 2024 Annual Report.* AI Index Steering Committee, Institute for Human-Centered AI, Stanford University. https://aiindex.stanford.edu

[32] World Economic Forum. (2022). *Responsible AI and ESG: Aligning business governance with AI ethics.* https://www.weforum.org/agenda/2022/06/responsible-ai-business-governance/

5. Australia's RAI related regulatory updates

a. Australia's 8 AI Ethics Key Principles[33]

Australia has developed a set of 8 ethical principles to guide RAI.

	AI Ethics Key Principle	Description
1.	**Key Principle:** Human, societal and environmental wellbeing **Explanation:** AI systems should benefit individuals, society and the environment. **Categories Impacted:** • Extraction of Data • Extraction of Minerals • Extraction of Labour	**Aspiration:** AI systems should promote societal benefit and environmental sustainability. This includes minimising environmental risks such as greenhouse gas emissions, energy overuse, and resource depletion. Mechanisms like renewable energy adoption, efficient hardware design, and regular environmental impact assessments help mitigate these effects. **Sources of Risk:** The training and use of large-scale AI systems can require massive amounts of processing power, memory, networking, storage, and other resources—collectively known as 'AI compute'—which can have significant environmental footprints from energy to water use, GHG emissions. [34] **Solutions:** The environmental costs of AI, such as energy-intensive data processing, can be mitigated through renewable energy sources and efficient hardware design. For instance, Google has implemented carbon-intelligent computing to reduce emissions in its data centres.
2.	**Key Principle:** Human-centred values **Explanation:** AI systems should respect human rights, diversity, and the autonomy of individuals. **Categories Impacted:** • Extraction of Data • Extraction of Labour	**Aspiration:** AI systems must uphold human dignity, rights, and values. They should reflect diversity and inclusivity, protecting vulnerable groups and respecting cultural and linguistic differences. Mechanisms include risk mitigation through regular assessments and providing remedies when harm occurs. **Sources of Risk:** AI systems may inadvertently perpetuate biases present in training data, leading to decisions that do not respect human rights, diversity, or individual autonomy. **Solutions:** Implementing AI impact assessments to measure real and potential harm, ensuring diverse and representative training data, and establishing processes for redress and remedy when harm occurs can help uphold human-centered values.

---

[33] Australian Government, Department of Industry, Science and Resources. (2019). *Australia's AI Ethics Principles*. https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles

[34] OECD. (2022). Measuring the environmental impacts of artificial intelligence compute and applications (p.15). https://www.oecd-ilibrary.org/science-and-technology/measuring-the-environmental-impacts-of-artificial-intelligence-compute-and-applications_7babf571-en

3.	**Key Principle:** Fairness	**Aspiration:** AI systems must be free from bias and discriminatory practices. They should ensure fairness in their design, data use, and outcomes. This involves validating datasets and algorithms to address systemic biases and support equitable access and opportunities. **Sources of Risk:** Flawed data or algorithmic design may result in discriminatory practices, leading to exclusion and inequality, particularly affecting gender equality and accessibility. **Solutions:** Conducting thorough validation of datasets and algorithms to identify and mitigate biases, and fostering inclusion and equality through equitable access to AI technologies can promote fairness.
	Explanation: AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups.	
	Categories Impacted: • Extraction of Data • Extraction of Labour	
4.	**Key Principle:** Privacy protection and security	**Aspiration:** AI systems must safeguard user data and privacy, ensuring compliance with legal standards. This includes preventing unauthorised access, maintaining secure data storage, and mitigating risks of information breaches through robust cybersecurity measures. **Sources of Risk:** AI systems are susceptible to privacy breaches, unauthorised data access, and security threats, including IP infringement and malicious use of AI. **Solutions:** Implementing robust management systems, policies, and tools for identifying and responding to threats, enforcing security measures, and ensuring compliance with privacy laws and regulations can protect privacy and security.
	Explanation: AI systems should respect and uphold privacy rights and data protection, and ensure the security of data.	
	Categories Impacted: • Extraction of Data	
5.	**Principle:** Reliability and safety	**Aspiration:** AI systems must perform consistently and safely under all operating conditions. Regular testing, validation, and monitoring are essential to ensure the system operates as intended, adapts to new challenges, and avoids harm to users. **Sources of Risk:** Inconsistent performance, inability to handle abnormal conditions, and susceptibility to manipulation can compromise the reliability and safety of AI systems. **Solutions:** Conducting regular performance testing, validation, and monitoring, and designing systems capable of adapting to new conditions without harming users can enhance reliability and safety.
	Explanation: AI systems should reliably operate in accordance with their intended purpose.	
	Categories Impacted: • Extraction of Data	

6.	**Key Principle:** Transparency and explainability	**Aspiration:** AI systems should provide clear and understandable information to stakeholders. Transparency mechanisms should explain how decisions are made, the data used, and the algorithms applied. Accessible documentation and independent audits enhance stakeholder trust.
	Explanation: 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.	**Sources of Risk:** Lack of transparency can lead to misunderstandings, mistrust, and inability to identify or rectify errors in AI systems. **Solutions:** Ensuring transparency of the intended outcome, data, algorithms, and logic applied in AI systems, and extending transparency to all stakeholders, including developers, end-users, and regulators, can improve explainability.
	Categories Impacted: • Extraction of Data • Extraction of Labour	
7.	**Key Principle:** Contestability	**Aspiration:** AI systems must allow individuals and groups to challenge decisions or outcomes that significantly impact them. This involves implementing timely processes for review, ensuring decisions are fair, and providing accessible paths for redress.
	Explanation: 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.	**Sources of Risk:** Without mechanisms for contestability, individuals may be adversely affected by AI decisions without recourse, leading to unfair outcomes and lack of trust. **Solutions:** Establishing clear and accessible processes for individuals to challenge AI decisions, and ensuring timely and fair reviews of contested outcomes can uphold contestability.
	Categories Impacted: • Extraction of Data • Extraction of Minerals • Extraction of Labour	
8	**Key Principle:** Accountability	**Aspiration:** AI systems require clear accountability structures that define roles and responsibilities across the AI lifecycle. Human oversight must govern the system's deployment, ensuring ethical compliance and adherence to legal obligations.
	Explanation: 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.	**Sources of Risk:** Ambiguity in roles and responsibilities can lead to ethical lapses, legal non-compliance, and lack of accountability in AI system outcomes. **Solutions:** Clearly defining the roles and responsibilities of all stakeholders involved in the AI lifecycle, ensuring human oversight, and maintaining compliance with legal requirements and ethical principles can enhance accountability.
	Categories Impacted: • Extraction of Data • Extraction of Minerals • Extraction of Labour	

b. Australian Government's 10 Guardrails (announced Sept 2024)

Australia has also announced a proposal for 10 mandatory guardrails for the safe and responsible use of AI in Australia, that complement the 8 ethical principles.[35]

The 10 Guardrails[36]	
1.	Establish, implement and publish an accountability process including governance, internal capability and a strategy for regulatory compliance.
2.	Establish and implement a risk management process to identify and mitigate risks.
3.	Protect AI systems and implement data governance measures to manage data quality and provenance.
4.	Test AI models and systems to evaluate model performance and monitor the system once deployed.
5.	Enable human control or intervention in an AI system to achieve meaningful human oversight across the life cycle.
6.	Inform end-users regarding AI-enabled decisions, interactions with AI and AI-generated content.
7.	Establish processes for people impacted by AI systems to challenge use or outcomes.
8.	Be transparent with other organisations across the AI supply chain about data, models and systems to help them effectively address risks.
9.	Keep and maintain records to allow third parties to assess compliance with guardrails.
10.	Engage your stakeholders and evaluate their needs and circumstances, with a focus on safety, diversity, inclusion and fairness.[37]

Aim

The mandatory guardrails aim to regulate 'high-risk' AI settings, and they will apply to both developers and deployers. Relevant considerations include the frequency of use, the level of risk of the AI setting, and the dynamics between different combinations (e.g., low use case + high-risk AI setting, high use case + low-risk AI setting). The definition and scope of a 'high-risk' AI setting is awaiting confirmation pending further consultation, whilst a voluntary model comprising standards aligning with the mandatory model has been proposed for 'low risk' AI settings.[38]

---

[35] The guardrails currently remain a set of aspirational principles, with no proposed enforcement mechanism to date.

[36] Australian Government, Department of Industry, Science and Resources. (2024). *The 10 Guardrails*. https://www.industry.gov.au/publications/voluntary-ai-safety-standard/10-guardrails

[37] An inclusive approach to stakeholder engagement should involve regular consultations with underrepresented communities, interdisciplinary experts, and public interest groups to ensure equitable AI outcomes.

[38] The standards for low-risk and high-risk AI settings are the same (the same guardrails), the difference is that it's proposed to be mandatory for high-risk settings and voluntary for low-risk settings.

Coverage

The guardrails aim to cover three main aspects of RAI governance:

- **Testing** – during development and in deployment to ensure systems perform as intended and meet appropriate performance metrics
- **Transparency** – about how AI products are developed and used with end-users, other actors in the AI supply chain and relevant authorities
- **Accountability** – for governing and managing risks associated with AI systems.

Enforcement

The regulatory options (i.e., how to enforce the mandatory guardrails) currently await further consultation with three potential options on the table, each with its pros and cons.

- Option 1: adapt existing regulatory frameworks on sector specific basis (Sector Specific)
- Option 2: adapt regulatory frameworks through framework legislation (Whole of economy)
- Option 3: introduce new standalone AI Act (Whole of economy)

c. New OAIC AI Guidance (published Oct 2024)

The Office of the Australian Information Commissioner (OAIC) has published new guidelines to clarify how Australian privacy laws apply to AI and set regulatory expectations.[39]

The current Privacy Commissioner Carly Kind emphasised the importance of strong AI governance and privacy safeguards to build public trust. The OAIC's focus includes addressing privacy risks from AI and advocating for privacy reforms, such as ensuring the fair use of personal information.

- **Guideline 1: Guidance on Privacy and the Use of Commercially Available AI Products.**

---

[39] Office of the Australian Information Commissioner. (2024). *New AI guidance: Privacy and AI governance in Australia.* https://www.oaic.gov.au

Helps businesses comply with privacy rules when using AI products.

Key Points

1. **Privacy Obligations:** Privacy rules apply to all personal data entered into or generated by AI systems, including inferred data (like deepfakes) related to identifiable individuals.

2. **Due Diligence and Product Selection:** Organisations should conduct thorough due diligence when adopting AI products. This includes verifying suitability, embedding human oversight, assessing privacy/security risks, and maintaining ongoing monitoring.

3. **Transparency and Policy Updates:** Privacy policies should clearly disclose AI use, and public-facing AI tools must be identified. Organisations should establish policies for AI use to support transparency and privacy governance.

4. **Sensitive Information Handling:** Best practices discourage using personal, especially sensitive, information in public AI tools due to complex privacy risks. If used, consent or reasonable expectation criteria must be met, particularly for secondary uses of personal data.

5. **Accuracy and Privacy by Design:** AI systems often produce inaccurate results. Under APP 10, entities must ensure data accuracy and use transparency tools (e.g., disclaimers). A 'privacy by design' approach, including a Privacy Impact Assessment, is recommended.

- **Guideline 2:** **Guidance on Privacy and Developing and Training Generative AI Models**

   Provides guidance to AI developers handling personal data for training generative AI models.

   Key Points

   1. **Accuracy and Risk Management**: Developers should ensure the accuracy of generative AI models by using high-quality datasets and conducting thorough testing, especially given the heightened risks associated with AI applications. They may also need to use disclaimers for high-privacy-risk uses.

2. **Data Usage Legality**: Just because data is publicly available does not mean it can be legally used for training AI models. Developers must assess whether the data contains personal information and adhere to privacy regulations, potentially requiring the deletion of certain information.

3. **Sensitive Information Handling**: Developers must be cautious with sensitive information, which typically requires consent for collection. This includes photos or recordings that may contain identifiable data, which cannot be collected without proper consent.

4. **Secondary Use of Personal Information**: If developers plan to use personal data already in their possession for training AI models—when that was not the original purpose—they must consider privacy obligations. They should ensure that the secondary use aligns with what individuals reasonably expected when their data was collected.

5. **Consent and Opt-Out Mechanisms**: If a secondary use of data for AI-related purposes does not meet expectations or relate closely to the original purpose, developers should seek explicit consent from individuals or provide a meaningful opt-out option to mitigate regulatory risks.

d. RAI and ESG for Practitioners (published Oct 2024)

The Australian Government published a **practical guide for ESG practitioners regarding AI use** in October 2024.[40] The guide provides a comprehensive introduction to how AI intersects with ESG initiatives. It outlines both the potential benefits and risks of using AI in ESG contexts, emphasising AI can accelerate sustainability outcomes, such as reducing carbon emissions and enhancing accessibility, while also posing significant risks related to bias, privacy, and environmental pressures. The guide offers practical examples, including AI applications in accessibility, financial protection, and energy efficiency, and stresses the importance of responsible AI governance, transparency, and

---

[40] Australian Government, Department of Industry, Science and Resources. (2024). *Practical guide for ESG practitioners regarding AI use.* https://www.industry.gov.au

partnerships in addressing these challenges. It encourages ESG practitioners to integrate AI into their strategies by leveraging existing frameworks, collaborating with AI developers, and using tools like the AI Impact Navigator to measure AI's societal and environmental impacts.

e. RAI & ESG for Investors (published April 2024)[41]
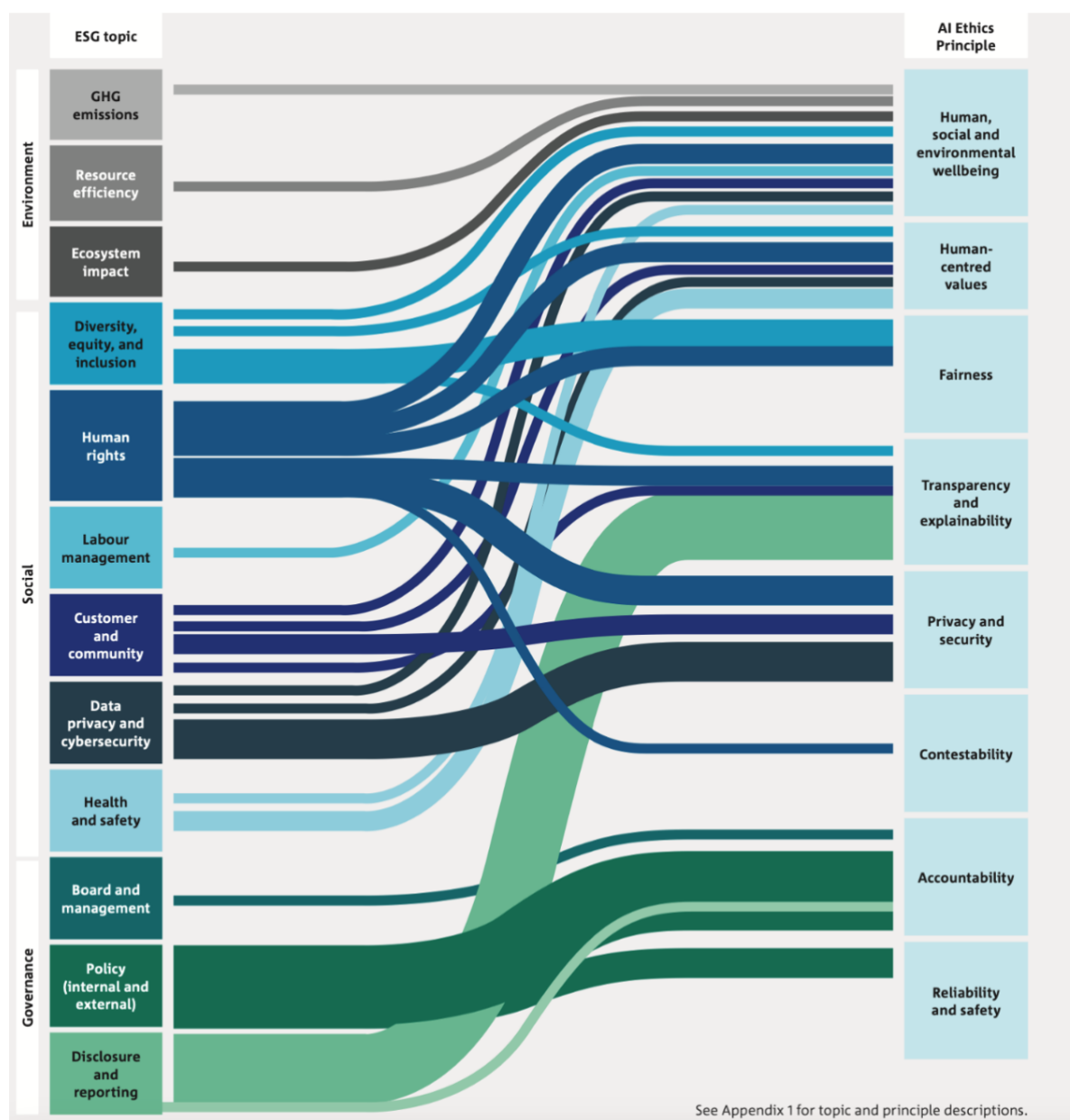
Source: CSIRO & Alphinity Investment Management, *The Intersection of Responsible AI and ESG: A Framework for Investors*, April 2024.

---

[41] CSIRO & Alphinity Investment Management. (2024). *The intersection of responsible AI and ESG: A framework for investors*. Commonwealth Scientific and Industrial Research Organisation. https://www.csiro.au/-/media/D61/Responsible-AI/Alphinity/Responsible-AI-and-ESG.pdf

	ESG topic	Description	Example AI applications
Environmental	**GHG emissions**	Training and running AI models takes a significant amount of energy. However, AI can also reduce emissions through asset optimisation, automation and operational efficiency.	Digital twins and asset modelling improve operational efficiency and reduce fuel use. AI algorithms support the energy grid by predicting demand and supply fluctuations. This can optimise energy flow, balance the grid, prevent outages and ensure consistent energy supply.
	Resource efficiency	AI can play a role in optimising resource efficiency in operations and across the supply chain. Depending on the industry this can help reduce energy, land and water consumption.	Predictive maintenance using AI-powered tools can optimise maintenance schedules. AI can optimise logistics, predict demand and improve quality control
	Ecosystem impact	AI can play a role in tackling environmental challenges. It can bring big data into the picture to monitor and address key ecosystem threats and opportunities across issues such as deforestation, soil health and pollution.	AI-enabled satellite imagery and geospatial mapping can monitor environment impacts and land use change. AI-enabled early warning systems can detect hazards such as bushfires or pollution events in real-time, allowing for timely intervention.
Social	**Diversity, equity and inclusion (DEI)**	AI can perpetuate existing biases or even introduce new forms of discrimination. AI can also support inclusion when trained on up-to-date, high-quality and diverse datasets.	Financial services use AI in application processes and to help make credit decisions. AI in healthcare allows clinicians to adopt data -driven diagnosis and deliver services remotely. AI can support inclusion, such as hearing and visual aids for people with disability or automated machinery.
	Human rights	Using AI for surveillance, weapons, to spread misinformation, and to reduce access for select groups can breach human rights. AI can also help to address issues such as modern slavery through greater supply chain transparency and information sharing, and increasing the use of robotics to automate low-value and unsafe tasks.	AI-driven surveillance and monitoring such as facial recognition and other image analysis tools. AI can use supply chain datasets to generate meaningful insights about modern slavery and human rights risks. Automation is integrated into the production of goods that rely on low-skilled, repetitive and manual human labour.
	Labour management	Using AI to automate repetitive or manual tasks in workforces can boost employee satisfaction, address labour shortages and improve productivity outcomes. However, it could also result in job losses, particularly affecting those in lower-paid roles who already face challenges with financial security.	Automation changes the employment landscape and reduces manual, repetitive and mundane tasks. Wearable technology can collect employee data, monitor activities and support safety and productivity outcomes. AI-integrated hiring supports employee selection.
	Customer and community	AI efficacy, security, accuracy, accountability, transparency and reliability pose reputational risks for companies. Companies that safely adopt AI can improve product quality, expand market reach, better service stakeholders such as customers, and benefit from recognised leadership related to AI opportunities.	Product development and innovation from selling AI tools or using AI to power existing processes. Customer service such as chatbots and virtual assistants can give 24/7 support and manage routine enquiries. AI insights to model and calculate insurance prices.
	Data privacy and cybersecurity	The use of big data to power AI increases risks related to data privacy, fraud and security, and consent. On the other hand, AI can support fraud detection and help to support cybersecurity by detecting threats and performing predictive analysis.	AI use cases require data and digitalisation, exposing companies to privacy and cybersecurity risks. AI systems in health research use particularly sensitive and personal datasets. AI algorithms can detect fraudulent activity in financial or consumer sectors.

ESG topic		Description	Example AI applications
Health and Safety		AI systems can recognise trends and correlations for potential hazards, allowing organisation to minimise high-severity injuries and fatalities. This also comes with a risk of automated systems failing and causing injury.	AI-enabled sensing devices detect unsafe practices or working conditions that could lead to accidents or fatalities. Automation can reduce physical strain of manual labour, especially from repetitive tasks.
Governance	Board and Management	Leadership awareness and capability play an important role in an organisation's success in an AI-enabled world	
	Policy (internal and external)	An RAI policy can be an early indicator of AI leadership and can build trust by serving as an explicit commitment to ethical AI practices.	
	Disclosure and reporting	Although ESG disclosures are improving, RAI disclosures remain nascent. Good quality disclosures are important to maintain a strong social licence to operate, prepare for future reporting requirements and ensure transparency with stakeholders.	

Source: CSIRO & Alphinity Investment Management, *The Intersection of Responsible AI and ESG: A Framework for Investors*, April 2024.

f. Generative AI Practice Note and Judicial Guidelines (published Nov 2024)

Practice Note and Guidelines for the use of Generative AI (Gen AI) in the NSW legal system, effective **3 February 2025**.[42]

Key points

1. **Commencement and Scope**: The Practice Note SC Gen 23 applies to all proceedings and covers both closed-source and open-source Gen AI.

   a. Closed-source Gen AI – proprietary tools developed by private organisations, with inaccessible source code and training data. E.g., ChatGPT, Google Bard, and CoCounsel Core.

   b. Open-source Gen AI – have publicly available source code and training data and enables collaborative development, customisation, and

---

[42] Supreme Court of New South Wales. (2024). *Practice Note SC Gen 23: Use of Generative Artificial Intelligence*. https://www.supremecourt.justice.nsw.gov.au/Documents/Practice%20and%20Procedure/Practice%20Notes/general/current/PN_Generative_AI_21112024.pdf

adaptation for specific needs. E.g., LLaMA and Hugging Face Transformers.

2. **Definition of Gen AI:** AI that creates content (text, images, sounds) based on training data.
    a. e.g., ChatGPT, Google Bard, Westlaw Precision, Co-Counsel...
    b. Excludes basic tools for spelling, grammar correction, formatting, and traditional search engines.

3. **Risks of Gen AI:**
    a. Hallucinations – the generation of seemingly credible but inaccurate or fabricated responses, such as false citations and invented legal or secondary references.
    b. Outdated, biased, or irrelevant data
    c. Confidentiality, privacy, or copyright concerns.
    d. Training programs for legal and technical professionals on assessing Gen AI risks, such as fabricated outputs or hallucinations, will be essential for effective governance.

4. **Acceptable Use**: Gen AI can assist with tasks like drafting documents and summarising information but *must not be used for generating affidavits, witness statements, or expert reports* without prior court approval.

5. **Prohibitions of Gen AI Use:**
    a. Gen AI must not be used for inputting confidential or legally restricted material, such as suppressed evidence, subpoenaed materials, or information subject to non-publication or suppression orders.
    b. Generating affidavits, witness statements, or expert reports without prior approval.

6. **Expert Reports**:
    a. Gen AI use requires *court leave* and *disclosure* of specific details (tool, benefit, process)
    b. Must keep records of AI usage.

7. **Written Submissions:**
    a. Any use of Gen AI must be disclosed.

b. All legal and factual references must be manually and independently verified for accuracy, relevance, and existence of citations and references

8. **Judicial Guidelines**:

   a. Judges are prohibited from using Gen AI for drafting judgements of analysing evidence.

   b. If used for secondary research, judges must verify the accuracy and completeness of the output.

   c. Associates or researchers must disclose Gen AI use and verify outputs.

9. **Transparency and Disclosure**: Legal practitioners, unrepresented parties, and judicial staff must disclose any use of Gen AI and verify its output.

10. **Review and Updates**:

    a. Disclosure of Gen AI usage is mandatory in certain documents.

    b. Violations, such as undisclosed usage, will be scrutinised.

    c. Both the Practice Note and Guidelines will be periodically reviewed due to the evolving nature of Gen AI technology.

6. Conclusion

In Summary

Australia has made significant steps towards developing a framework for RAI with the 2024 Guardrails and 8 AI Ethics Principles, demonstrating a commitment to ethical, transparent, and accountable AI governance. However, Australia still faces key challenges in sustainability, enforceability, and organisational AI maturity.

While Australia performs well in governance and ethical AI frameworks, its reliance on voluntary guidelines contrasts with the more stringent regulatory approaches seen in the EU, South Korea, and Canada. Australia's global ranking in Responsible AI suggests a solid foundation, but areas such as technical capabilities, investment in AI research, and sustainability measures require attention.

The Role of New Non-Mandatory Guidance in Strengthening RAI

However, importantly Australia has introduced several non-mandatory initiatives to support responsible AI adoption, which—if widely implemented—could significantly enhance governance, industry compliance, and ethical AI maturity:

(a) New OAIC AI Guidance (October 2024): This provides privacy safeguards and regulatory expectations for AI developers and businesses. This guidance can help mitigate risks related to data misuse, AI-generated misinformation, and privacy breaches, by preserving privacy, data governance, and accountability.

(b) RAI & ESG Guidance for Investors (April 2024): This initiative bridges the gap between AI ethics and environmental, social, and governance (ESG) considerations, helping businesses integrate sustainable AI practices into their investment strategies. Greater ESG integration could lead to more responsible AI adoption, particularly in sectors where AI affects labour markets and environmental sustainability.

(c) Generative AI Practice Note & Judicial Guidelines (November 2024): These initiatives provide practical governance for the use of AI usage in legal and administrative settings, addressing risks like 'hallucinations', misinformation, and privacy concerns in AI-generated content. The legal sector's adoption of this
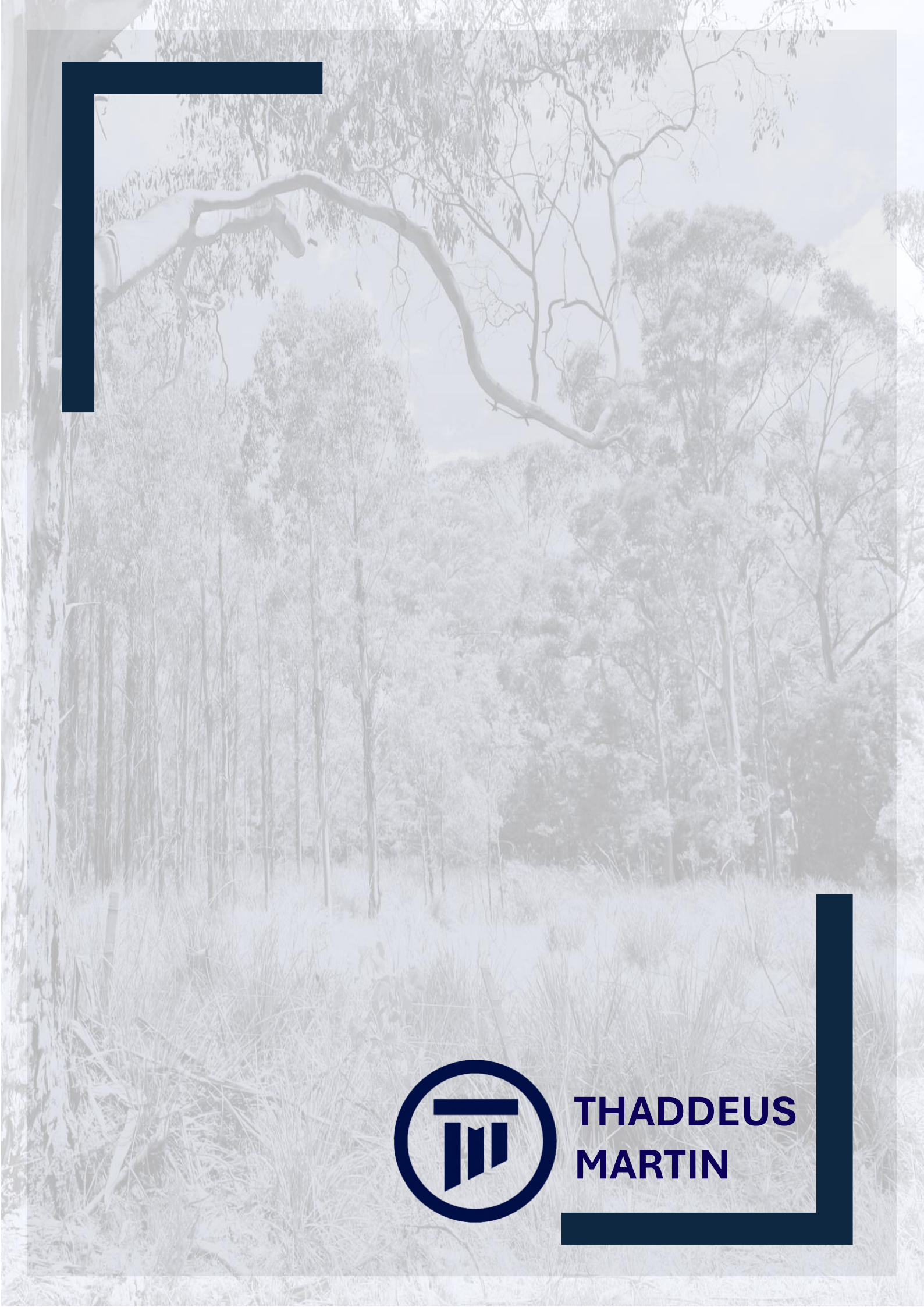
legal Practice Note and Guidelines set a positive precedent for broader AI accountability and transparency across other industries.

Future Directions

These non-mandatory measures offer valuable guidance, but their impact depends on their voluntary adoption by Australian business and industry. Australia should explore pathways to progressively transition key principles into enforceable regulations to maximise their effectiveness, particularly for high-risk AI applications.

Therefore, it is prudent to recommend that Australia's RAI strategy should evolve from principles-based guidance to enforceable legislation, whilst ensuring AI governance remains both ethical and adaptable to emerging challenges. Strengthening regulatory enforcement, expanding AI education and public sector capabilities, and ensuring Indigenous data sovereignty and sustainability integration will be critical for positioning Australia as a global leader in Responsible AI.

Australia has the opportunity to set a global precedent for balancing innovation with ethical AI governance, by addressing these gaps and aligning its frameworks with international best practices.

THADDEUS
MARTIN