The intersection of Responsible AI and ESG: A Framework for Investors

Suggested citation

Alphinity Investment Management (Alphinity) and Commonwealth Scientific and Industrial Research Organisation (CSIRO), The intersection of Responsible AI and ESG: A Framework for Investors, CSIRO, 2024.

Copyright

CSIRO disclaimer

Alphinity disclaimer

Project team

Jessica Cairns
Head of ESG and Sustainability
Alphinity Investment Management

Moana Nottage
ESG and Sustainability Analyst
Alphinity Investment Management

Mary Manning
Portfolio Manager
Alphinity Investment Management

Qinghua Lu
Responsible AI Science Team Leader
CSIRO's Data61

Sunny Lee
AI Governance and Architecture Lead
CSIRO's Data61

Harsha Perera
Postdoctoral Researcher
CSIRO's Data61

Sarah Kaur
Senior Design Thinking Practitioner
CSIRO's Data61

Judy Slatyer
Lead, Responsible AI@Scale Think Tank
National AI Centre

Edited by **Shae Lalor** AE, The Write Path

Contents

This report

This report presents the insights and outcomes developed through a collaborative partnership between CSIRO's Data61 and Alphinity Investment Management. It is intended to be used by equity investors who want to assess the environmental, social and governance (ESG) implications of the design, development and deployment of Artificial Intelligence (AI). It can also be used as a guide for listed companies and other stakeholders that are considering how best to integrate efforts in Responsible AI (RAI).

The ESG-AI Framework presented in this report is the main outcome of the partnership. A set of templates have also been developed to help investors implement the framework.

How to use this report

We encourage investors to:

- read the 10 key insights from the company engagements and research
- understand Australia's AI Ethics Principles
- follow the ESG-AI framework's assessment steps (1 to 3, as needed)
- use the spreadsheet templates provided under a Creative Commons licence.

Scope

This report and framework are grounded in Australia's AI Ethics Principles and draw from CSIRO's question bank and metric catalogue. It can be used by investors to assess the integration of RAI for companies across all sectors. The insights and framework have been developed using information from large, listed companies; however, the concepts can apply to companies of all sizes and potentially extend to the unlisted space as well.

This report does not include consideration of sustainability frameworks or concepts such as the United Nations Sustainable Development Goals.

Company engagement and research

Thank you to the companies that contributed to this project. 28 companies participated in a research interview that informed the ESG-AI framework and insights. A further 25 companies, including 19 global and 6 Australian organisations, were analysed via desktop research.

Foreword

We're excited to present this essential guidance for investors on how to assess the responsible use of Artificial Intelligence (AI).

This **investor framework** has been developed based on an extensive 12-month research project – a partnership between CSIRO's Data61 and Alphinity Investment Management. Initial discussions related to sustainable technology early in 2022 ignited this project and identified the need to bridge the gap between ESG and Australia's AI Ethics Principles.

Why?

AI is a **rapidly growing force** impacting how companies build new markets, drive productivity improvements and enhance customer engagement. Yet they also need to know how to govern it responsibly in line with their strategies. Critically, investors need to understand the **key threats and opportunities** of AI usage and consider their implications on environmental, social and governance factors (ESG).

What?

The **transformative potential of AI** is already shaping how companies operate. Based on extensive research and engagement with Australian and globally listed companies, this report presents key insights, company case studies, and detailed guidance for implementing the framework. Metrics for responsible AI are not often shared publicly, making it challenging for investors to assess performance and understand what best practice looks like.

Who?

The partnership authors are strong advocates for responsible investment. Together, they bring a deep knowledge and authority on environmental, social, and governance (ESG) integration, responsible AI and sustainable investment. Their engagement with companies already responsibly navigating the complexities of AI provides leading insights.

How?

This framework not only delivers insightful research, but it provides practical tools for investors by operationalising Australia's 8 AI Ethics Principles. It is hoped the investment community – and companies more broadly – will embrace these tools as standard practice for responsible AI measurement.

The framework considers the work of others in this space, including initiatives sponsored through the **National AI Centre** to develop a framework for companies to report on the external impacts of the use of AI, and the **Responsible AI Think Tank**, which convenes thought leaders to inform the strategic priorities of industry adoption of AI at scale.

> We hope investors adopt this framework and become advocates for an improved AI future.
> We also encourage companies to integrate this framework into their management of AI threats and opportunities.

CSIRO's Data61 and Alphinity Investment Management

The commercialisation of AI brings with it a promising future

GROWTH MARKET
BY 2025: the market for AI software will reach **$135 billion. Market growth will double** from 14% in 2021 to 31% in 2025[1]

LISTED COMPANY INTEREST
Highest number of S&P 500 companies **citing 'AI'** on Q2 earnings calls in over 10 years[2]

TRANSFORMING OUR WORKPLACES AND WORKFORCES
2/3 occupations at risk of automation[5]

$600 billion of anticipated economic activity facing Gen AI disruption[3]

AUSTRALIA'S POTENTIAL
BY 2028: digital and AI tech worth **$315 billion**[1]

INVESTMENT OPPORTUNITY
2030: 7x amount invested annually in AI

IN 5 YEARS: double daily users[4]

CHANGING REGULATION
31 countries have passed AI legislation and 13 more are debating AI laws[5]

AI has the potential to make businesses more efficient, reduce costs, revolutionise business practices, improve employee experience, and generate revenue from new or enhanced products and services.

But for these opportunities to be realised, the governance, design, and application of AI needs to be undertaken in a responsible and ethical way.

Investors play a crucial role in shaping **Australia's Responsible AI (RAI) ecosystem**, yet there is a lack of established guidance for investors to navigate this rapidly evolving space.

> RAI is the practice of developing and using AI systems in a way that provides benefits to individuals, groups, and wider society, while minimising the risk of negative consequences.[6]

For investment and societal opportunities to be realised, the governance, design, and application of AI needs to be undertaken **responsibly, safely and ethically**.

While RAI **governance and implementation** is in its early days, overlaying RAI onto existing **ESG frameworks** offers a useful way for investors to effectively analyse threats and opportunities.

But existing ESG frameworks do not include metrics or measures related to RAI. Some frameworks cover issues which may be **linked** to AI, for example, data privacy and cyber security; however, there is no specific guidance for investors to analyse AI-related ESG threats or opportunities.

> Unlike traditional data models, dispersed AI systems may amplify errors and lead to decisions that cause harm to individuals, society and the environment. To add to this, the 'black box' nature of AI and the ability to self-govern complicates the process of ensuring its responsible conduct.

Managing the ethical implications of AI and safeguarding against regulatory and reputational risks unlocks the full potential of **AI-driven innovation**.

---

1   Artificial Intelligence Roadmap: Solving problems, growing the economy and improving our quality of life (CSIRO, 2019).
2   Highest number of S&P 500 companies citing AI on Q2 Earnings (Factset, 2023).
3   Generation AI: Ready or not, here we come! (Deloitte Australia, 2023).
4   Generative AI: A quarter of Australia's economy faces significant and imminent disruption (Deloitte Australia, 2023).
5   AI Regulation is Coming – What is the Likely Outcome? (Centre for Strategic and International Studies, 2023).
6   Qinghua Lu, Liming Zhu, Jon Whittle, Xiwei Xu, Responsible AI: Best Practices for Creating Trustworthy AI Systems, Addison Wesley Professional, 7 December 2023.

Project overview

Early in 2023, the Alphinity and CSIRO teams agreed to co-develop a Responsible AI framework for investors. The teams worked closely to engage with Australian and globally listed companies and conduct desktop research. The ESG-AI Framework presented in this report is the main outcome of the partnership. A set of templates have also been developed to help investors implement the framework.

Methodology

The project team engaged with 28 listed companies to:

> **understand**
the state of play when it comes to AI uptake

> **identify**
good practice implementation of RAI governance, strategy and risk management

> **gain**
an understanding of company practices for those actively considering RAI

> **develop**
a framework for investors to assess RAI, building on CSIRO's RAI research, Australia's AI Ethics Principles and existing ESG foundations

US 18%

Europe 11%

Asia 11%

x28
INTERVIEWS
AROUND
THE WORLD

Australia

60%

Real estate 3.6%

Industrial 10.7%

Energy 10.7%

Consumer discretionary 10.7%

Information technology 21.4%

8 SECTORS

Communication 10.7%

Materials 7.1%

Financial 25%

Although this framework has been developed for investors, it is also hoped this research will serve as a guiding tool for company reporting on RAI. The project started in February 2023.

2023				2024	
FEB	APR	AUG	OCT	DEC	JAN	MAR

Planning	**Phase 1** **Direct company engagement** Conduct interviews with target companies	**Phase 2** **Framework design** Assess interviews, design framework, grey literature research, input expert opinion	**Phase 3** **Framework development and reporting** Confirm framework, develop templates, finalise engagement insights
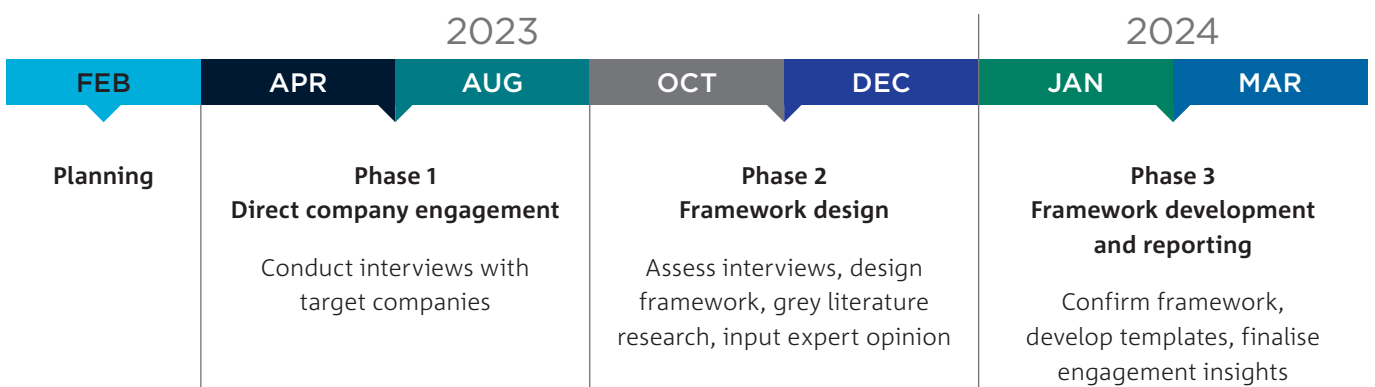
Defining AI

AI is a **convergence of technologies**, including computing power, scalability, networking, connected devices and data. It leverages computers and data to perform tasks traditionally requiring human intelligence.

OECD defines an AI system as "a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their level of autonomy and adaptiveness after deployment"[7].

While defining AI is important, there are several terms that intertwine with AI in our day-to-day life. For instance:

- **Narrow AI** (e.g. facial recognition) represents AI use in today's world and essentially focuses on a particular problem.[8]

- **General-purpose AI** is a type of AI system that addresses a broad range of tasks and uses, often referred as 'next step of future AI' (e.g. Natural Language Understanding [NLU]).

- **Generative AI** (GenAI) is a branch of AI that develops generative models with the capability of learning to produce content such as images, text, and other media with similar properties as their training data.[9]

For the purpose of this framework, we have considered all types of AI and the way companies embrace **responsible AI (RAI) practices**. According to Lu et al., RAI is 'the practice of developing and using AI systems in a way that provides benefits to individuals, groups, and wider society, while minimising the risk of negative consequences.'[10] Similarly, the Bletchley Declaration defines AI safety as 'for the good of all, AI should be designed, developed, deployed, and used, in a manner that is safe, in such a way as to be human-centric, trustworthy and responsible.'[11]

We also define RAI using *Australia's AI Ethics Principles*. These principles were endorsed by the Australian Government in 2019: 'It's part of the Australian Government's commitment to make Australia a global leader in responsible and inclusive AI. For Australia to realise the immense potential of AI we need to be able to trust it is safe, secure and reliable.'[12]

---

7 OECD, https://oecd.ai/en/wonk/ai-system-definition-update.

8 ISO, *ISO-IEC-22989 Artificial intelligence concepts and terminology*, ISO, 2022.

9 A Vassilev, A Oprea, A Fordyce and H Anderson, 'Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations', *NIST Computer Security Resource Center*, January 2024.

10 Qinghua Lu, Liming Zhu, Jon Whittle, Xiwei Xu, Responsible AI: Best Practices for Creating Trustworthy AI Systems, Addison Wesley Professional, 7 December 2023.
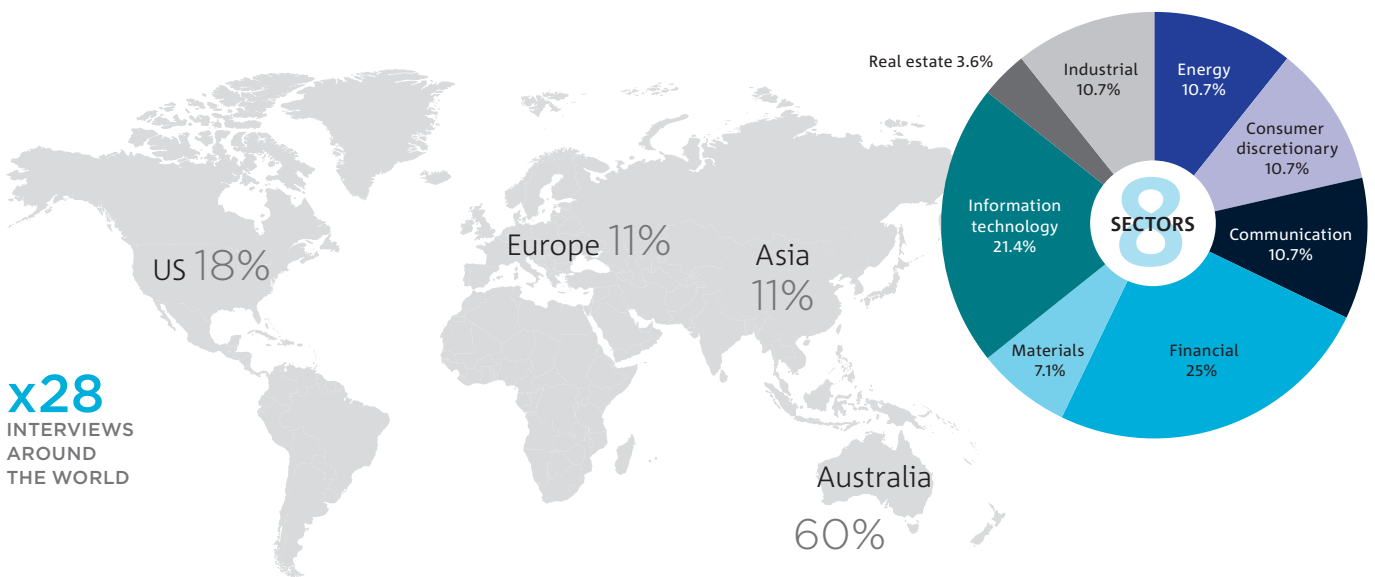
11 UK Government, *Policy paper: The Bletchley Declaration by Countries Attending the AI Safety Summit, 1–2 November 2023*, published 1 November 2023.

12 Australian Government Department of Industry, Science and Resources, *Australia's Artificial Intelligence Ethics Framework*, Australian Government, 7 November 2019.

The intersection between Responsible AI and ESG

We define RAI in line with the 8 AI ethics principles.[13] Standard ESG concepts are a way to operationalise the principles. This illustration maps 12 environmental, social and governance topics against the principles.

ESG considers both threats and opportunities, and this balance is equally important when thinking about the benefits versus harm of AI.
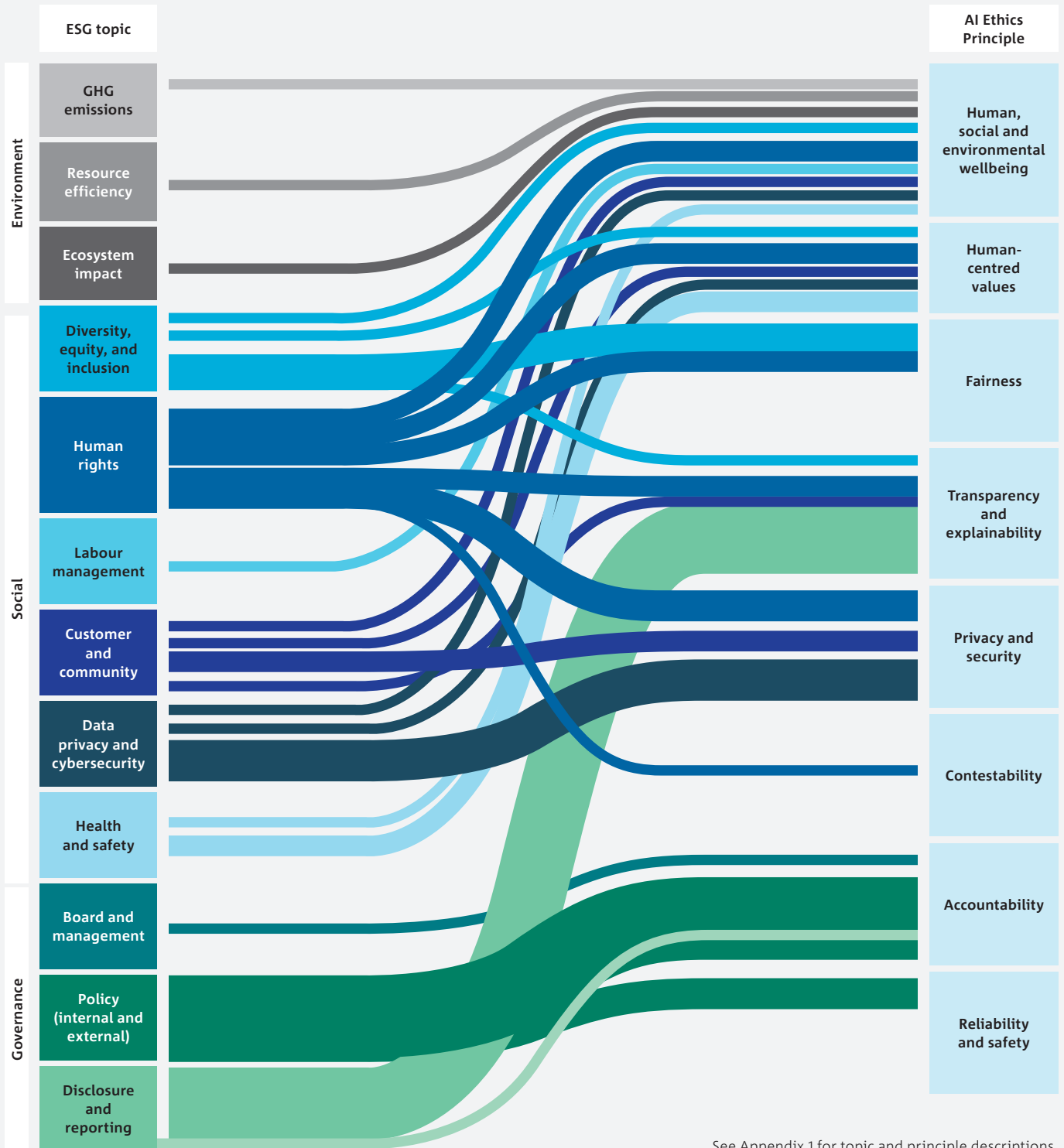
See Appendix 1 for topic and principle descriptions.

---

13 Australian Government Department of Industry, Science and Resources, *Australia's AI Ethics Principles*, Australian Government, n.d.

AI investment landscape

Recent advances in AI are transforming the way we work and live. AI is also changing the investment landscape by creating many new investment opportunities and risks. Like other technology-led industrial revolutions, the widespread commercialisation and dissemination of AI will create both winners and losers from an investment perspective.

There are 3 waves of AI investment opportunities. Figure 1 illustrates these waves over time.

AI OPPORTUNITY SET

Wave 3
AI cost saving opportunities and innovation

Wave 2
AI revenue opportunities (other key sectors)

Wave 1
AI revenue opportunities (TMT sectors)

TIME

Figure 1: Investment opportunities in AI

The **first wave** is already well underway. It is dominated by companies that have direct revenue exposure to AI-related products and services as well as 'picks and shovel' stocks that provide the tools, platforms and infrastructure required to driv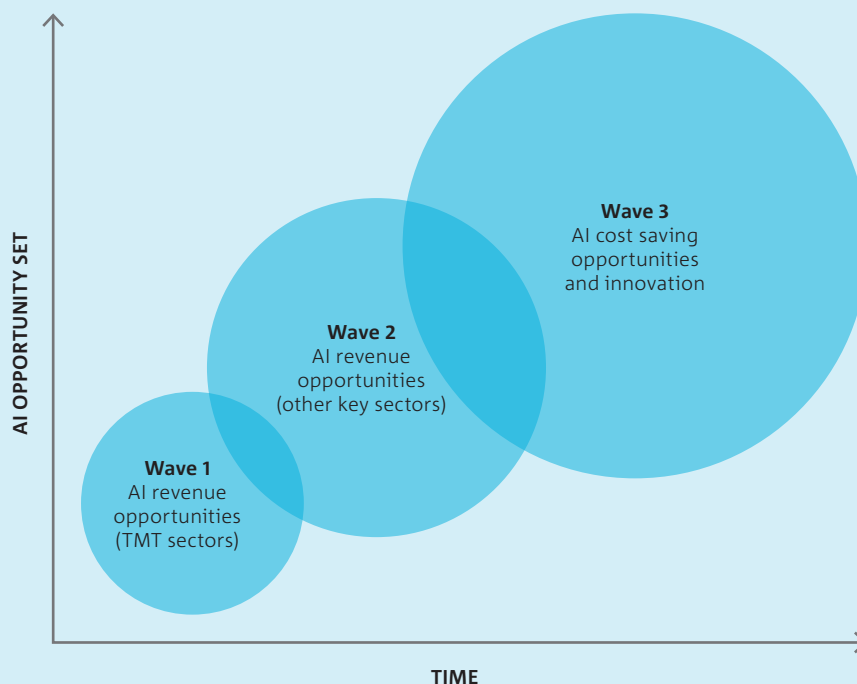e success in an AI-enabled world. These companies are primarily in the Technology, Media and Telecom (TMT) sectors and are primarily listed in the United States.

The most obvious example in this category is Nvidia, which designs and supplies the Graphics Processing Units (GPUs) required to train Large Language Models (LLMs).

The **second wave** of AI winners will emerge across all sectors and geographies. This wave is focused on companies that can use AI to increase their revenue and earnings, even if they do not sell AI products or services directly.

A good example in this wave is Airbnb. Airbnb is planning to use AI to transition from an alternative accommodation business to an ultimate concierge business. Revenue and earnings will potentially increase as AI is used to provide hyper-personalised experiences for its users and expand the company's reach from property rentals to multiple travel and accommodation verticals. In the Australian market, a good example in this category is Woolworths, which is using its Quantium data analytics business to increase advertising and sales opportunities.

The **third wave** is the most comprehensive but will also take the most time. It refers to the AI opportunities that will arise for companies that use AI to reduce costs and improve productivity. For example, McKinsey estimates that the deployment of AI and other technologies could provide the global economy with an annual productivity boost of 0.5% in 2023 to 3.4% in 2040, depending on the rate of automation adoption. McKinsey estimates that GenAI alone can add up to 0.6% of this growth.[14] The report illustrates that the estimated **productivity** benefits in Wave 3 will significantly outweigh the direct and indirect **revenue** benefits of AI by a wide margin. Further opportunities are also anticipated from significant innovation and breakthroughs driven by AI.

---

14 McKinsey & Company, 'The economic potential of generative AI: The next productivity frontier', *McKinsey Digital*, 14 June 2023.

At a company level, cost savings and productivity gains will accrue disproportionately to companies with high levels of low value-added tasks, easily automated operations and use cases that can easily be disrupted by AI. Examples include banks and insurers, health care companies, telecommunications providers, materials companies and retailers.

Read more in the company case studies on page 19.

It is critical to highlight that investment in new and emerging AI opportunities comes with **significant risk**. Technological risk, execution risk, valuation risk, regulatory risk and ESG risks are all important factors when assessing investments and outcomes.

This framework is designed to help investors address the ESG risks associated with AI while also focusing on potential opportunities.

Company insights

These insights were developed using information gathered through interviews with 28 Australian and globally listed companies. They are supported by desktop research of a further 25 companies, including 6 Australian and 19 global organisations.

1

Only a small percentage of companies publicly disclose their RAI policies

40% of interviewed companies had internal RAI policies, yet only 10% shared these publicly. This trend is reinforced by our desktop review and findings from the World Benchmarking Index, revealing that only 19 out of 200 evaluated companies announced their AI principles.[15]

34% of invited companies declined to be interviewed. The companies that declined cited reasons such as their own AI maturity levels or concern about market sensitivity.

Behind the scenes, companies are actively exploring AI opportunities. 62% of interviewed companies were either starting, or had implemented, an AI strategy for the business.

Commonwealth Bank has recently published its inaugural AI policy.

2

Global equities are at the forefront of AI implementation

Extensive AI resourcing among global companies. Many of the global companies interviewed were advanced in the implementation of AI strategies.

Shell began developing business use cases for what we would now call machine learning in the 1970s and 1980s, leveraging advanced statistical methods for scenario planning and product testing. Around 2013, the company recognised to integrate AI into operational processes through software and started developing a cloud architecture.

> 'Almost all of Shell's assets are connected to a common data platform, and the number of AI use cases stretches into the hundreds.'
> — Shell[16]

MercadoLibre responsibly incorporates AI across its operations, leveraging machine learning to boost e-commerce sales, detect fraud at Mercado Pago, enhance logistics efficiency, and optimise lending and credit analysis at Mercado Credito.

> 'Ethics and user trust are paramount in our approach to AI.'
> — MercadoLibre

Many Australian companies are only just beginning to consider opportunities for AI. Of those interviewed, the Australian banks are the most advanced.

3

Employee engagement is essential to deliver AI-related opportunities

A culture of curiosity is important to identify cross-functional opportunities. Successful AI implementation requires input from both technical and non-technical staff. For example, engineers and consultants need to generate AI-related ideas for developers so the business needs can be met effectively. These types of partnerships are particularly crucial in industries like industrial and mining, where technology adoption has traditionally been limited.

The **Shell** AI community was put in place in 2013 and now has over 11,000 people involved. The original purpose was to build awareness around AI. Now, the community are leading much of the change in AI across the business. Having a structure of engaged employees has helped Shell to pivot quickly when needed.

AI training and awareness supports a strong risk management culture. Staff can build confidence around the new technologies, and feel comfortable speaking up about risks, issues and new opportunities.

---

15 World Benchmarking Alliance, *Augmenting Ethical AI: 2023 Progress Report on the Collective Impact Coalition for Digital Inclusion*, September 2023.

16 The quoted materials have been sourced from company interviews. Consent was obtained for the use of all quotes and case studies in this report.

4

Strengthening Board and leadership capability in AI, technology and ethics

Directors need tech know-how to navigate AI. Given the competitive landscape for experienced AI directors, alternative approaches such as training and raising awareness among existing Board members become essential but are yet to be fully explored. Companies with technology expertise are better placed to expand knowledge appropriately in the AI space. **MercadoLibre** has very strong technology and AI experience on the Board of Directors. **ANZ** also identified this as a need a number of years ago and added new Directors with the right experience.

Surprisingly varied AI reporting to Board. While numerous companies report AI opportunities and uptake to their Boards, this is ad hoc. Notably, the approach lacks consistency compared to other material ESG topics such as climate change or health and safety, which can be standing agenda items.

> Of the interviewed companies, 42% had at least one Director with strong capability in tech/AI (or two with some capability) or evidence that the company is focused on enhancing awareness and capability through training.

5

RAI governance is best embedded within existing systems and processes

Ethics and values should guide AI decisions. A phrase we heard time and again was the integration of corporate values into RAI practices. This ensures the company does not damage its social licence, customer trust, data privacy, or influence other material ESG topics.

Cross-disciplinary governance. This involves implementing governance structures that involve representatives from various disciplines to analyse risks and make informed decisions about AI strategy aligned with business objectives. For example, **Wesfarmers** does this well by having a group that includes leaders from different verticals of the company. They work together to decide on risks and what projects to focus on, and to find new opportunities.

Need for defined RAI responsibility and sensitive use cases. Microsoft is recognised for its robust RAI governance structure and leading RAI framework, particularly in explicitly referencing sensitive use cases.

> 'We want to have digital systems that reflect our corporate values. A technology may tick the boxes and still not feel right for our culture. That's why we include Woodside values in our RAI framework.' – Woodside Energy

> 'Any system that incorporates AI technology and meets the definition of a sensitive use case must be reviewed by the Office of Responsible AI.' – Microsoft

> 'Our 'do no harm' principle is fundamental when we introduce new technology.' – Westpac

6

Strong track record in ESG performance is an indicator of confidence for investors

Companies prioritising stakeholder impact. Companies that carefully consider how their actions affect people, their reputation, and how they're seen by society are likely to approach new technologies like AI with the same care. These companies generally have well-respected Boards, robust disclosures and ESG commitments, and are likely to implement AI responsibly and in a measured way.

ESG ambitions are a proxy for good AI management. Because AI is evolving so rapidly, good leadership on existing topics like cyber, diversity and employee engagement is a proxy that the impact of AI will also be considered thoughtfully. **Keysight**, a well-known technology and ESG leader, provides a 'safe sandbox' environment conducive to exploring AI technologies cautiously.

> 'It's important we have a safe sandbox in which to explore AI tools.' – Keysight Technologies

One of the core framework pillars centres on good governance. This deliberate choice underscores the significance of governance in shaping responsible and impactful AI deployment, ensuring a thorough evaluation of AI practices. **Mirvac** exemplifies good governance practices by deeply considering these factors and taking a risk-based approach to AI implementation, which mirrors its ESG leadership and strong performance record.

> 'Mirvac uses a risk–reward matrix. Our approach to AI is context specific and risk aware. The whole area is full of unknown unknowns, but it will clearly impact most industries and we need to get in front of it.' – Mirvac

7

A balanced view of threats and opportunities is needed to mitigate harm and leverage AI benefits

Overly cautious approach hinders progress. Many companies express concerns about the potential negative impacts on their reputation, consumer trust, and regulatory consequences. While caution is understandable, this shouldn't stifle innovation and the potential for productivity gains. For example, some companies restricted employees from using AI tools such as ChatGPT, while others took an educational stance.

Engagements revealed that companies with good overall governance structures were more likely to balance threats and opportunities brought by AI, therefore displaying a healthy curiosity into this technology. Conversely, companies with weak overall governance were unlikely to show leadership characteristics with respect to the development and implementation of RAI and could therefore limit the opportunity brought by AI. **Therefore, it's crucial to establish effective AI guardrails to ensure safe AI deployment.**

Importance of targets in RAI strategy. Accenture stands out for its significant investment in RAI and the publication of a toolkit to facilitate seamless integration of RAI into its operations. It is among the few companies with publicly stated AI targets.

> 'We are investing US$3 billion into AI and doubling the AI workforce from 40k to 80k to accelerate client reinvention.' – Accenture

8

Companies are using different strategies for navigating and managing RAI risk, but supply chain management can be overlooked

Lack of awareness among non-AI developers. Many of the companies interviewed hadn't considered managing risks through procurement. Addressing this gap is crucial, especially for sectors that are less tech savvy, to establish an ethical AI ecosystem that goes beyond individual organisational boundaries.

> 'Always insource strategy, governance and risk, and hold responsibility for the outcomes.' – ANZ

RAI needs to be considered in strategic partnerships and with suppliers. Commonwealth Bank's collaboration with H2O.ai to develop AI solutions highlights the significance of strategic partnerships in navigating AI complexities. It also clearly shows there is a distribution of risks and responsibilities, which purchasers must consider seriously and disclose performance against.

9

Most companies are investing in AI, but RAI policies and reporting are still being developed

Despite considerable RAI activity, it is not always reflected in external reporting. Many companies are actively involved in RAI initiatives, yet they may not be emphasising this to the market. Some companies fail to mention AI in their risk statements, strategic pillars and annual reports, despite expressing enthusiasm about it in discussions and making significant investment into exploring the technology.

Discrepancy in public policies and implementation. While some companies have established RAI frameworks, there were repeated cases where there was little evidence of actual implementation.

> The desktop review found that all 25 companies had announced a recent AI initiative, but only 52% highlight AI as a key opportunity in recent annual reports.

10

Data privacy is a key ESG issue, but other topics are still important and may be overlooked

Data privacy is the most common concern. During the interviews, data privacy and cyber security were consistently identified as the issues most material to AI.

Limited attention to human rights. Human rights and modern slavery were not identified as concerns in the interviews. With one of the core AI ethics principles focusing on human rights, this topic remains critical, yet underexplored in the AI space.

Concerns about AI used for safety benefits was surprising. Some companies have trialled AI-driven safety measures, but employees expressed concerns about privacy and surveillance, particularly about biometrics and monitoring through wearables and cameras that would otherwise offer significant safety benefits.

Repeatedly heard that AI will augment and not replace jobs. Companies had a strong stance that AI will not replace employment, but instead lead to increased productivity and streamlining of mundane tasks.

> 'Managing data privacy and governance is a big priority' – Transurban

The following table presents a list of ESG issues and example AI applications which may have a positive or negative impact.

Table 1: Examples of ESG issues impacted by different AI applications

ESG	AI APPLICATIONS
Diversity, equity and inclusion	**Financial services** use AI in application processes and to assist credit decisions. **AI in healthcare** enables clinicians to adopt data-driven diagnosis and deliver services remotely. **AI to support inclusion**, such as hearing and visual aids for people with disability or automated machinery.
Human rights	**AI-driven surveillance and monitoring** such as facial recognition and other image analysis tools. **Supply chain datasets** can be utilised by AI 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	**Automation** changes the employment landscape and reduces manual, repetitive and mundane tasks. **Wearable technology** can collect employee data, monitor activities, and enable safety and productivity outcomes. **AI-integrated hiring** supports employee selection.
Customer and community	**Product development and innovation** from selling AI tools or using AI to power existing processes. **Customer service** such as chatbots and virtual assistants can provide 24/7 support and manage routine enquiries. AI insights to **model and calculate insurance prices**.
Data privacy and cybersecurity	**AI use cases require data and digitalisation**, exposing companies to privacy and cybersecurity risks. Use of **AI systems in health research** use particularly sensitive and personal datasets. AI algorithms can **detect fraudulent activity** in financial or consumer sectors.
Health and safety	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.
GHG emissions	**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	**Predictive maintenance** using AI-powered tools can optimise maintenance schedules. AI can **optimise logistics**, predict demand and improve quality control.
Ecosystem impact	AI-enabled **satellite imagery and geospatial mapping** can monitor environmental 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.

The ESG-AI Investor Framework

This framework has 3 components underpinned by 12 ESG topics that are relevant to AI.

Foundations of this framework

- **Leading standard in RAI.** We anticipate the number of companies that will address all requirements of this framework will be limited initially; however, investors can use the measures and metrics to drive enhanced disclosure and outcomes over the longer term.

- **Threat and opportunity view.** Investors and companies need to focus on realising the opportunity as well as managing the threat.

- **Flexibility and materiality assessment.** Components can be used individually or all together. Factors that determine materiality have been embedded to guide investors on what is important and to support flexibility.

- **Underpinned by ESG.** Designed to bridge the gap between existing ESG theory and the AI ethics principles to enable RAI assessment.

- **Questions and metrics.** Informed by existing CSIRO research and RAI question bank[17] and metric catalogue.

- **Established AI ethics principles.** Guide investors and companies around the ethical considerations of AI.

- **Regulatory flags.** Risk categories from the EU AI Act are integrated within the use case analysis to flag potential compliance, transparency and management requirements.

Explore our ESG-AI mapping in the section on page 5.

Using the framework

This framework has 3 components. These can be used together or individually depending on investor preference:

- **AI use case analysis:** 27 material AI use cases for 9 different industries offer a threat- and opportunity-based view on different AI technologies for investors.

- **RAI governance indicators:** Aspects such as Board oversight, public commitments and implementation inform investors of a company's position on RAI.

- **RAI deep dive:** Guiding questions and metrics around Australia's 8 AI Ethics Principles to complete detailed analysis and support enhanced AI disclosure.

Given RAI is still emerging, we expect that at least initially, investors will need to engage with companies to collect all the information required to answer the framework questions in full. Over time, as awareness of RAI and disclosure improves, and more investors use this framework to engage with companies, we hope the level of disclosure related to RAI will improve.

---

17 SU Lee, H Perera, B Xia, Y Liu, Q Lu, L Zhu, O Salvado and J Whittle, *QB4AIRA: Question Bank for AI Risk Assessment*, Data61 CSIRO, 11 July 2023.w

Framework at a glance

	Higher-level threat and opportunity-based analysis Best for screening or analysis across a large group of companies Completed using public disclosures or engagement		Deeper analysis of management practices Requires engagement with company experts to complete Can support enhanced AI disclosures
	Step 1 **AI use case analysis**	**Step 2** **RAI governance indicators**	**Step 3** **RAI deep dive**
	Identify companies exposed to material AI use cases and determine next steps based on environmental and social impact areas.	Complete high-level analysis across 10 RAI governance indicators to determine the overall strength of a company's management approach.	Facilitate detailed analysis and engagement with company management on AI governance and RAI practices.
Overview	**Materiality assessment for 27 key AI use cases across 9 key sectors.** Materiality is determined by 3 factors: • regulatory risk • environmental and social impacts (positive/negative) • impact scope (industry or systemic).	**10 indicators that can be used to assess the overall commitment, accountability and measurement of RAI.** The governance indicators are split across 4 categories: • Board oversight • RAI commitment • RAI implementation • RAI metrics.	**Deep dive questions and indicators to assess company performance against Australia's AI Ethics Principles.** There are 42 sub-questions, which can be rolled up to one leading question per principle. Select questions by features such as: • organisational type (AI purchaser and/or developer) • AI system category • ESG topics.
Template	The template is pre-populated with material use cases by sector, with an assessment against the 3 materiality factors.	The template includes 10 indicators with assessment guidance.	The template has been set up with assessment questions, detailed descriptions, guide metrics and ESG alignment.
Application	Screen groups of companies or sectors to identify the relevance of high-, medium- and low-risk use cases. Outputs may inform further analysis using Step 2 and/or Step 3, investment case considerations and/or stewardship priorities.	Evaluate the suitability of company governance processes to manage RAI risks. Some indicators can be assessed using public disclosures. To assess all 10 indicators, a brief engagement with an Investor Relations representative or similar is recommended.	Analyse high-risk companies (i.e. based on exposure to material AI use cases — see Step 1) and identify gaps across RAI management practices. This is detailed analysis and requires deep desktop review and targeted engagement with company experts. Recommended for companies that are investing heavily into AI and are exposed to material use cases.
Outputs	High-, medium- and low-risk use cases by sector Environmental and social impact summaries per material use case	RAI governance score	Ethics principle assessment: • principle question score (unacceptable, weak, acceptable, and strong) • principle sub-question score (0–5-point scale)

See how the framework is applied in the example on page 17.

Step 1: AI use case analysis
Materiality assessment for 27 key AI use cases across 9 key sectors

Use this step to identify companies exposed to high-impact AI use cases and determine next steps based on regulatory risk, environmental and social consideration and impact scope.

This component has been developed by identifying potential materiality factors through a comprehensive review of academic literature, regulatory guidelines such as the EU AI Act, and industry AI frameworks, including the OECD[18] and Microsoft RAI framework.[19] These insights were further enriched by engagement with companies.

Investors can use these 3 key materiality factors to produce a high-, medium- or low-risk level for the AI use case.

- **Regulatory risk:** Europe is now a global standard-setter in trustworthy AI. The adoption of the EU AI Act marks a new phase in the global race on AI policy. Under the risk-based approach of the Act, AI systems are divided into 5 categories (4 categories from the Act and 1 new category) according to the associated societal risk (Appendix 2).

- **Environmental and social impacts:** Each use case carries both positive and negative implications, which we have aligned with 9 environmental and social topics. The governance impacts are company-specific and sector-agnostic and are therefore covered in other parts of the framework.

- **Impact scope:** It is crucial to consider whether a use case may affect the industry or have systemic implications that might cause harm, trigger regulatory changes, hinder AI exploration, or positively disrupt areas of the economy.

See Appendix 2 for risk assessment definitions and descriptions. Templates are available online: csiro.au/RAI-ESG-Report.

---

18 OECD, *OECD Framework for the Classification of AI Systems*, OECD Publishing, 22 February 2022.

19 Microsoft, *Microsoft Responsible AI Impact Assessment Guide,* Microsoft, June 2022.

Step 2: RAI governance indicators
10 indicators to assess the overall commitment, accountability and measurement of RAI

Use this step to assess a company's maturity in RAI governance. Investors should pay attention to Board oversight, public commitment to RAI, implementation of the RAI policy (or similar commitment) and how it discloses RAI measures (see Table 2). This component has been developed using standard governance frameworks, public company disclosures and insights from the interviews.

Investors can use these 10 indicators to specify key RAI disclosures that companies should prioritise and report in the short-term.

Detail on the RAI governance indicators can be found in Appendix 3. The template is available online.

Table 2: RAI governance indicators

CATEGORY	INDICATOR	DESCRIPTION
Board oversight	1. Board accountability	RAI is explicitly mentioned as part of the responsibility of the Board or a relevant Board subcommittee (e.g. risk committee or ESG committee). Board receives structured RAI reporting at least once per year but more frequently as needed.
	2. Board capability	At least one Director with strong technology-related experience.
RAI commitment	3. Public RAI policy	Policy should align with relevant industry standards (e.g. ISO/IEC 42001, Australia's AI Ethics Principles). The RAI policy should include consideration of ethics, company values, testing and transparency.
	4. Sensitive use cases	Sensitive, high-risk use cases (such as facial recognition) are addressed as part of the RAI policy. Sensitive use cases require additional oversight and approval.
	5. RAI target	RAI policy or commitment is supported with clear targets (e.g. % of workforce trained, reduction in RAI incidents).
RAI implementation	6. Dedicated RAI responsibility	RAI oversight can be dedicated, or part of another role or function.
	7. Employee awareness	Specific program in place to increase employee awareness of AI, alongside relevant ethical and ESG considerations.
	8. System integration	RAI policy is integrated throughout existing business processes, including risk management, product development, procurement and ESG.
	9. AI incidents	Issues and incidents related to RAI are tracked and reported internally.
RAI metrics	10. RAI metrics	RAI metrics (such as the use of AI) associated with the policy are identified and reported externally to stakeholders.

SCORE: X/10

🔴 Weak 0–3   🟡 Moderate 4–7   🟢 Strong 8–10

Step 3: RAI deep dive
Deep dive questions and indicators to assess company performance against Australia's AI Ethics Principles

Use this step to facilitate detailed analysis and engagement with company management on AI governance and RAI practices. AI ethics principles encompass key values and guidelines that address RAI development and deployment. Use this assessment for a systematic evaluation of fairness, transparency, accountability, privacy, and more, contributing to a holistic understanding of how well a company adheres to ethical standards in its AI practices.

This assessment has been developed based on the CSIRO RAI question bank and metric catalogue, which draws on insights and standards from key regulatory bodies, standard organisations and stakeholder groups, such as:

- EU AI Act
- NIST AI Risk Management Framework[20]
- ISO AI Standard (ISO/IEC 42001)[21]
- other industry AI risk frameworks.

Investors can use this component to undertake research on specific ESG concerns or use cases. This step in the framework is **flexible**. Tailor the questions based on your ESG interests or by material principles. There are 42 assessment questions distributed among the 8 ethics principles with 27 specific indicators (see Table 3). For example, you can:

- adopt the 8 leading principle questions in company engagement and RAI analysis
- conduct a deep dive at the principle level by applying filters
- conduct a deep dive at the ESG topic level by applying filters and utilising the sub-questions
- complete a full assessment where companies have exposure to high-risk AI use cases and score poorly against the RAI governance indicators.

See Appendix 4 for example questions and metrics. Templates are available online.

Table 3: Key components of the deep dive assessment

COMPONENT	DESCRIPTION	COUNT (#)
Principle question	A dedicated question designed to assess a company's overarching adherence to that principle. Sub-questions help to elaborate on each principle.	8
Principle sub-question	A series of questions per principle. Designed to elicit information and insights about a specific aspect of a company's AI practices.	42
Guide metrics	A series of metrics assigned to each principle sub question. Designed to encourage companies in measuring and disclosing specific aspects of their RAI operations, practices or performance, and comply with regulations (e.g. EU AI Act).	43

20 NIST, *AI Risk Management Framework*, NIST Information Technology Laboratory, NIST, 26 January 2023.
21 ISO, *ISO/IEC 42001, Artificial Intelligence Management System*, ISO, 2023.
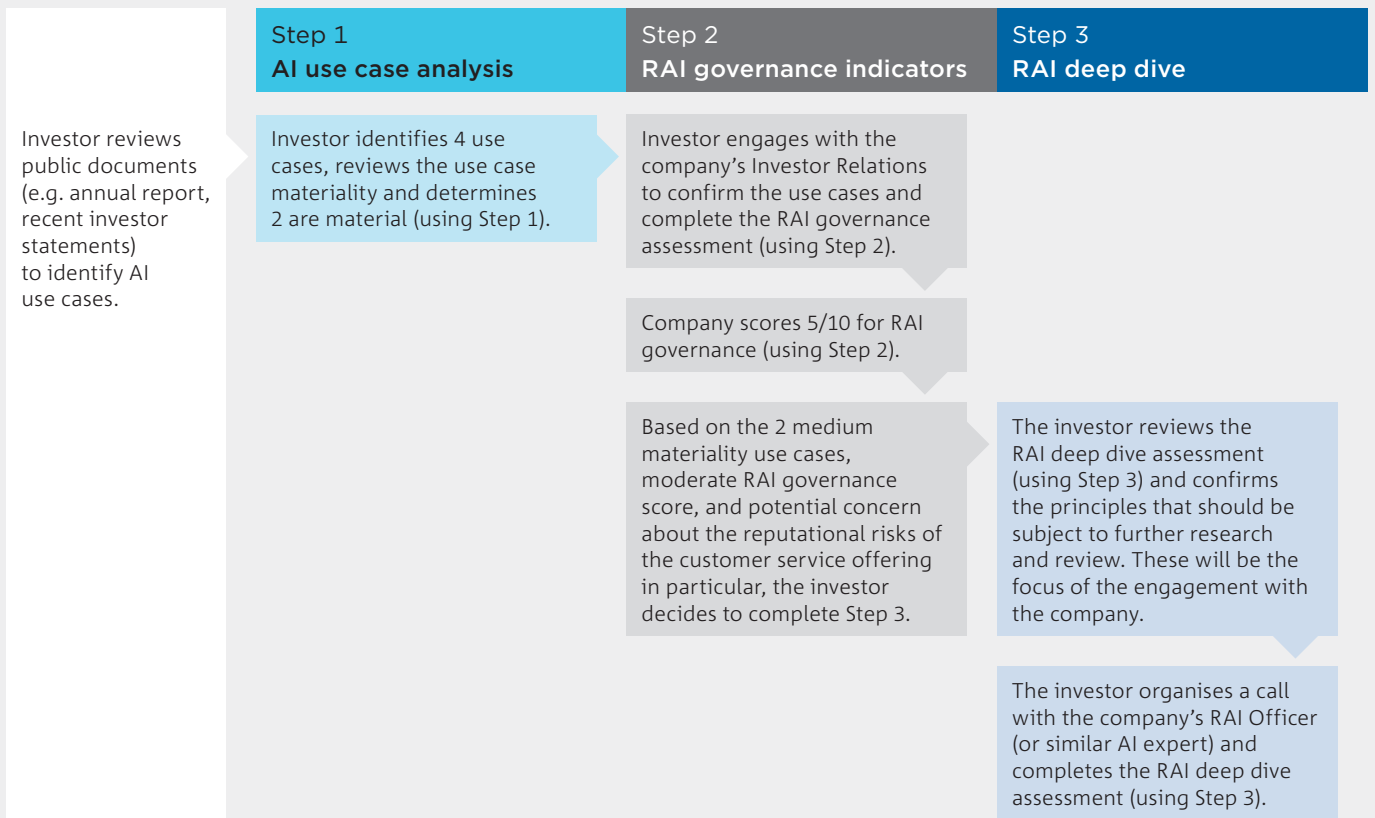
Applying the framework to assess the AI risks for a consumer company

This is an example of how the full 3-part framework can be used.

An equity investor would like to understand the ESG threats and opportunities associated with an Australian consumer company that is exploring the use of AI for its marketing and customer service. The company already uses AI for supply chain management, floor design, stock management and for an internal chatbot.

Assessment process

	Step 1 **AI use case analysis**	**Step 2** **RAI governance indicators**	**Step 3** **RAI deep dive**
Investor reviews public documents (e.g. annual report, recent investor statements) to identify AI use cases.	Investor identifies 4 use cases, reviews the use case materiality and determines 2 are material (using Step 1).	Investor engages with the company's Investor Relations to confirm the use cases and complete the RAI governance assessment (using Step 2).	
		Company scores 5/10 for RAI governance (using Step 2).	
		Based on the 2 medium materiality use cases, moderate RAI governance score, and potential concern about the reputational risks of the customer service offering in particular, the investor decides to complete Step 3.	The investor reviews the RAI deep dive assessment (using Step 3) and confirms the principles that should be subject to further research and review. These will be the focus of the engagement with the company.
			The investor organises a call with the company's RAI Officer (or similar AI expert) and completes the RAI deep dive assessment (using Step 3).

Assessment summary

Use case materiality: There are 2 medium-risk use cases that are currently used or planning to be adopted. These use cases present both threats and opportunities for the business.

FACTORS	USE CASE 1 SUPPLY CHAIN TRACEABILITY	USE CASE 2 CUSTOMER OFFERING
Regulatory risk	Medium	Medium
Environmental and social impacts	High	Medium
Impact scope	Industry	Industry
Future uptake	High	High

Internal chatbot, floor design and stock management are not considered material use cases.

Governance indicators: The company scored 5/10 against the governance indicators. Importantly, the internal RAI policy states high-risk use cases, such as facial recognition for surveillance in stores, are banned.

INDICATOR	ADDRESSED
1. Board accountability	YES
2. Board capability	YES
3. Public RAI policy	NO
4. Sensitive use cases	YES

INDICATOR	ADDRESSED
5. RAI target	NO
6. Dedicated RAI responsibility	YES
7. Employee awareness	NO

INDICATOR	ADDRESSED
8. System integration	YES
9. AI incidents	NO
10. RAI metrics	NO

RAI deep dive: Based on the outcomes of Step 1 and Step 2, the deep dive has been completed for four principles.
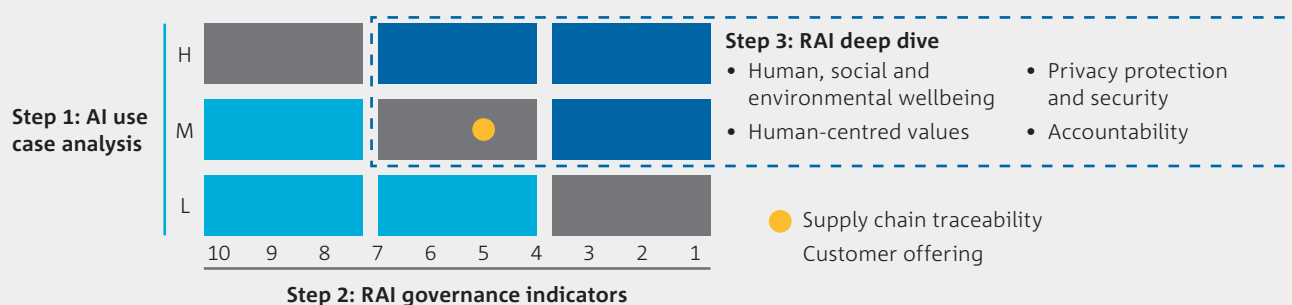
PRINCIPLE	SCORE	COMMENTS
Human, social and environmental wellbeing	Weak	The company has not completed environmental or social impact assessments to determine the potential impacts of the AI systems.
Human-centred values	Moderate	The company has a "human in the loop" position in its RAI policy. It has not included AI-related risks or impacts within its Modern Slavery Statement and it does not disclose diversity metrics of its AI team.
Privacy protection and security	Weak	The company has appropriate privacy and security policies and processes in place. There is a lack of evidence of regular and systematic auditing and reporting. There is also limited information about compliance with key privacy laws.
Accountability	Acceptable	The company identifies all relevant internal stakeholders and defines clear roles and responsibilities in relation to RAI. It operates an AI risk management system addressing serious incident cases that is incorporated in existing company risk management system. However, recordkeeping for the traceability of AI systems is not mentioned.

Assessment outcome

This company is using AI for 2 material use cases which are medium materiality. Using AI to support supply chain traceability presents significant efficiency and cost saving opportunity for the business. However, because it interacts with humans there are medium-level regulatory risks. The use of AI in marketing and customer service presents opportunities to expand reach, improve customer service, and improve marketing outcomes. However, there are privacy and security risks which may impact customer retention and the reputation of the business.

The company has addressed 5 out of 10 governance indicators. Considering the outcomes of the RAI governance assessment and RAI deep dive, four engagement objectives are recommended for investors to pursue:

- company to publish a RAI policy or framework
- company to complete a social impact assessment and confirm the benefits and constraints of using AI in supply chain and customer/marketing practices
- company to implement an employee engagement program on RAI to improve literacy in AI across the organisation
- company to improve disclosures related to privacy and security that specifically relate to audits and compliance with relevant privacy laws.

Step 1: AI use case analysis

Step 2: RAI governance indicators

Step 3: RAI deep dive
- Human, social and environmental wellbeing
- Human-centred values
- Privacy protection and security
- Accountability

🟡 Supply chain traceability
Customer offering

Company case studies

These case studies provide insights into how various companies that participated in the project are leading the way in responsible AI.

Accenture: Helping clients transition to an AI-enabled world

Accenture has positioned itself as a partner to guide clients through the transition to AI. In June 2023, Accenture announced a significant US$3 billion investment in its Data and AI practice to help clients rapidly and responsibly advance and use AI. This includes a target to double its AI talent from 40,000 to 80,000 professionals. This capital commitment is significant, and with it comes a responsibility to roll out AI ethically and closely consider well-known risks such as bias, accuracy and data security. From a governance point of view, Accenture is very clear that this is not one function's responsibility, and the ongoing focus is on implementing and upskilling responsible AI internally and with customers.

ANZ: Future proofing the business with a focus on accountability

ANZ have a set of Board endorsed ethics principles which apply to the design and application of AI systems. These principles are an extension of the company's corporate values and are used to evaluate different use cases for AI. ANZ is strongly focused on accountability and transparency and is aware of the reputational risks that come from using AI in customer facing products or services. It therefore aims to insource the development of AI for high-risk applications related to customers. ANZ are considering the longer-term risks and benefits of AI. It is planning out the future needs of its workforce and growing its technology roles accordingly. It has also identified impersonation and the use of AI in scams as a key concern in future.

Commonwealth Bank and H2O.ai: A successful partnership approach

The relationship between CBA and H2O.ai exemplifies the external partnership strategy. H2O.ai is a Silicon Valley-based company that operates a cloud-based Machine Learning (ML) platform called H2O AI Cloud. In 2021, CBA formed an exclusive partnership with H2O.ai and took a minority stake in the company. The partnership was designed to differentiate and advance CBA's capabilities across products, digital experiences and customer needs, and provide a platform for co-innovation between the 2 organisations. By jointly building AI solutions, they navigate the complexities of AI development while distributing risks and responsibilities.

Keysight: Effectively navigating regulatory uncertainty

Dr Mark Pierpoint, Vice President of Strategic Innovation and Partnerships at Keysight, serves on the Visiting Committee on Advanced Technology (VCAT) of the US National Institute of Standards and Technology (NIST). NIST is the agency of the US Department of Commerce that promotes innovation and industrial competitiveness, and it is currently focused on AI. Dr Pierpoint was part of the project engagement with Keysight and he and his team guide the company's technology investments and partnerships to develop future knowledge, capabilities and R&D. Engaging with government entities such as NIST and the AI Safety Institute Consortium (AISIC), and collaborating with key stakeholders both internally and externally, can help companies navigate an uncertain regulatory environment.

MercadoLibre: A Responsible AI leader in emerging markets

Mercado Libre (MELI) is the largest e-commerce and fintech company in Latin America. AI is infused throughout the organisation. In addition to pursuing AI opportunities across business units, MELI also has a keen eye on risk management and is using a 'human-in-the-loop' approach to minimise potential risks. The company acknowledged that this approach may increase the time required to develop and deploy AI tools across the organisation; however, it is an important part of risk management. MELI is also a leader, both globally and in emerging markets, in AI-related disclosure. The company's Transparency Report provides stakeholders with an update on the responsible use of technology, including AI.

Microsoft: Ahead of the game in RAI

Microsoft's vision is to empower transformation and unlock access to AI technology globally. Microsoft offers significant disclosure on its RAI standards, which are best-in-class because it demonstrates to stakeholders the 'how' of RAI implementation. Beyond disclosure, the Microsoft governance model is structured to embed RAI at every level of the business, including an Office for Responsible AI, a Responsible AI Council and 150 AI Champions dispersed globally. Microsoft is also an enable of responsible AI and works closely with its customers to educate and empower around this issue. It provides clear examples of sensitive AI use cases that are subject to enhance due diligence and escalation.

Mirvac: Shaping urban landscape through Responsible AI integration

Mirvac has established internal AI principles focused on bias, fairness, accountability, transparency and data privacy. All AI use cases are evaluated using a risk-reward matrix which reflects Mirvac's risk-conscious approach to technology adoption. Mirvac have been using AI in Machine Learning applications as part of it's engineering design processes for many years and are currently exploring opportunities related to generative AI. Mirvac's are putting in place governance structures to manage the ethical and business risks associated with the wider uptake of AI across the organisation.

Shell: AI use cases and the energy transition

The breadth of AI application at Shell is impressive in both breadth and depth. Interesting examples include using AI to detect reoccurring data patterns ahead of a pump failure, using AI-driven optimisation algorithms to improve EV charging through Shell Recharge Solutions, using AI-based technology in deep sea exploration, leveraging AI to improve worker safety and rolling out AI-enabled digital twins for Shell assets. The total number of AI use cases for Shell stretches into the hundreds, covering all businesses in the value chain of the energy sector. Shell started working on what we would now refer to as AI in the late 1970s and 1980s in the form of advanced statistical methods for scenario planning and product testing, and the longevity of Shell's approach to AI means it has a significant head start in thinking about RAI. Given its domicile, Shell active in its preparation for compliance with the EU AI Act, and is an active member of the Responsible AI institute.

Transurban: A leader in driving Responsible AI Innovation in toll road operations

Transurban's approach to responsible AI and governance is leading. The company is investing in AI applications across its tolling infrastructure, in asset engineering, and for incident detection and response. It is most focused on using AI to improve it's business processes and increase efficiency. As always, Transurban is conscious of maintaining its social licence to operate, and as such it has applied its leading governance and risk management practices to the roll out of AI including putting in place a RAI assurance framework and dedicated Board reporting. In the past five years, the analytics and enterprise data teams have grown to over 20 people with 6 people dedicated to machine learning related projects. Transurban's combined effort to identify opportunities for AI, increase resourcing, and managing ethical concerns sets a leading precedent for the industry and positions it well to capitalise on positive outcomes.

Westpac: Focusing on human oversight and ethics

Westpac focuses on ethical AI integration and prioritising leadership awareness through workshops. Spearheaded by the Chief Technology Officer, AI initiatives are focused on improving productivity while ensuring human oversight in customer interactions. Recent tests showcase AI's potential to enhance productivity, with GenAI tools yielding a remarkable 46% improvement in the sample. The company has updated its AI principles to align with European standards, and has an AI Working Group to oversee AI decisions. Underpinning Westpac's robust risk management approach is the 'do no harm' principle. AI is integrated into technology road maps to effectively support business outcomes. Westpac also collaborates closely with Data61 to enhance trust in AI systems, while the implementation of a register for AI applications ensures transparency.

Woodside: Advanced employee engagement in digital transformation

Woodside is a leader in digital technology adoption in the energy sector. The company uses an RAI framework which was established remarkably early in 2018/2019. Its advanced employee engagement initiatives include 40 AI full-time equivalents (FTEs) and monthly workshops to foster continuous idea generation and opportunity exploration. Collaborating within the oil and gas peer network, Woodside emphasises the importance of human involvement in AI processes. It prioritises value partnerships with key providers like IBM. Woodside's focus on optimising chemical processing, LNG, SAT, and robotics demonstrates a commitment to RAI integration and technological advancement.

Appendix 1
ESG topics and AI ethics principles

ESG TOPIC	AI ETHICS PRINCIPLE
GHG emissions: A significant amount of energy is required to train and run AI models; however, AI can also reduce emissions through asset optimisation, automation and operational efficiency.	**Human, social and environmental wellbeing:** AI systems should benefit individuals, society and the environment.
Resource efficiency: AI can play a role in optimising resource efficiency within operations and across the supply chain. Depending on the industry this can help reduce energy, land and water consumption.	**Human-centred values:** AI systems should respect human rights, diversity, and the autonomy of individuals.
Ecosystem impact: AI can play a role in tackling environmental challenges, bringing big data into the picture to monitor and address key ecosystem threats and opportunities across issues such as deforestation, soil health and pollution.	**Fairness:** AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups.
Diversity, equity, and inclusion: 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.	**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.
Human rights: The use of 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 the use of robotics to automate low-value and unsafe tasks.	**Privacy and security:** AI systems should respect and uphold privacy rights and data protection, and ensure the security of data.
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.	**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.
Customer and community: AI efficacy, security, accuracy, accountability, transparency and reliability pose reputational risks for companies. Companies that safely implement AI can enhance product quality, expand market reach, better service stakeholders such as customers, and benefit from recognised leadership related to AI opportunities.	**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.
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.	**Privacy and security:** AI systems should respect and uphold privacy rights and data protection, and ensure the security of data.
Health and safety: AI systems can recognise trends and correlations for potential hazards, allowing organisations to minimise high-severity injuries and fatalities. This also comes with a risk of automated systems failing and causing injury.	**Reliability and safety:** AI systems should reliably operate in accordance with their intended purpose.
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.	

The left column is grouped into **ENVIRONMENT**, **SOCIAL**, and **GOVERNANCE** categories.

Appendix 2
Summary of AI risk categories

CATEGORY	DEFINITION	GUIDANCE
Unacceptable risk	Significant potential to manipulate persons, vulnerabilities of specific vulnerable groups (children, people with disability), AI-based social scoring for general purposes done by public authorities.	AI systems under this category will be prohibited, meaning their development, use and placing on the market will be banned.
High risk	Based on the intended purpose of the AI system, this use case may create a high risk to the health and safety or fundamental rights of natural persons. Examples: • Biometrics • Critical infrastructure • Credit scoring • Resume-scanning tool	High-risk AI use cases are obligated to adhere to regulatory requirements, encompassing a robust risk management system, a comprehensive quality management system covering the system, model, and data aspects, meticulous recordkeeping practices, and the creation of technical documents to ensure transparency.
Medium risk	This AI use case falls outside the categories of unacceptable risk and high risk. Instead, it is classified as an application interacting with humans.	The users should be informed of specific transparency obligations, enabling them to make informed decisions about their interactions with AI systems. Minimal transparency requirements should be met to empower users in deciding whether to continue using the application.
Low risk	Low-risk AI applications, excluding unacceptable/high- and medium-risk ones, demonstrate a lower level of potential harm. Examples: • AI-enabled video games • Spam filters	No specific guidance for low-risk AI applications.
Not determined	The risk associated with the AI use case has not been definitively assessed or categorised. This could be due to various factors such as insufficient information, complexity, or ambiguity regarding the use case's impact or potential risks.	Further analysis or clarification may be needed to determine the appropriate risk level.

Adapted from the EU AI Act.

Appendix 3
Responsible AI governance indicators

CATEGORY	INDICATOR	DESCRIPTION	ASSESSMENT GUIDANCE
Board oversight: The capability of directors in the digital and AI skillset is crucial in an era where companies must be guided effectively through the challenges and opportunities presented by new technology. The Board should pursue an understanding of AI capabilities and limitations, undergo training and receive structured reporting of RAI performance. RAI should be stated as an explicit responsibility of Board member(s) or a relevant Board subcommittee.	**1. Board accountability**	RAI is a responsibility of a Director or Board subcommittee Structured reporting on RAI to Board	RAI is explicitly mentioned as part of the responsibility of the Board or a relevant Board subcommittee (e.g. risk committee or ESG committee). Board receives structured RAI reporting at least once per year but more frequently as needed.
	2. Board capability	Specific AI and/or technology capability on the Board	At least one Director with strong technology-related experience.
RAI commitment: In a rapidly changing AI landscape, a public RAI policy is an early indicator of AI leadership and can build trust by serving as an explicit commitment. The policy should be supported by RAI targets and refer to ethical considerations and how AI-related decisions fit within corporate values. Having a policy on red lines or sensitive use cases is a proactive step for companies to communicate their stance on different AI technologies and identify the ones that do not align with values and/or risk appetite.	**3. Public RAI policy**	RAI policy or framework is in place and externally published	Policy should be aligned with relevant industry standards (e.g. NIST, Australia's AI Ethics Principles). The RAI policy should include consideration of ethics and company values.
	4. Sensitive use cases	Specific use cases are managed as part of overall RAI commitment	Sensitive, high-risk use cases (such as facial recognition) are addressed as part of the RAI policy. Sensitive use cases require additional oversight and approval.
	5. RAI target	RAI commitment driven by targets	RAI policy or commitment is supported with clear targets (e.g. % of workforce trained, reduction in RAI incidents).
RAI implementation: An AI Officer or similar AI-dedicated role is important to provide strategic guidance, ensure ethical and responsible AI practices and manage AI risks. An AI management committee can support a structured approach to overseeing AI initiatives by bringing together cross-functional expertise and ensuring integration across business units. We expect for RAI to be integrated through established business systems, and issues related to AI tracked and reported internally. An AI employee awareness program provides individuals with the knowledge and skills necessary to understand, develop and implement AI technologies ethically and safely. It also fosters a safe working environment where concerns or issues from this new technology can be raised.	**6. Dedicated RAI responsibility**	Designated individual or function that has oversight for RAI	RAI oversight can be dedicated, or part of another role or function.
	7. Employee awareness	AI employee awareness program	Specific program in place to increase employee awareness of AI, alongside relevant ethical and ESG considerations.
	8. System integration	RAI is integrated in business processes	RAI policy is integrated throughout existing business processes, including risk management, product development, procurement and ESG.
	9. AI incidents	RAI incidents are tracked and reported internally to management	Issues and incidents related to RAI are tracked and reported internally.
RAI metrics: Reporting on RAI measures are still in their infancy; however, an RAI policy or commitment should be supported with clear targets and a strategy to execute on the headline RAI commitment. RAI metrics should be linked to the policy and reported externally to stakeholders.	**10. RAI metrics**	Externally reported RAI metrics	RAI metrics associated with the policy are identified and reported externally to stakeholders.

Appendix 4
Responsible AI deep dive

PRINCIPLE	PRINCIPLE QUESTION	INDICATOR	EXAMPLE SUB QUESTIONS	EXAMPLE METRICS
Human, social, environmental wellbeing: AI systems should benefit individuals, society and the environment.	Are the company's AI systems assessed to have a net positive benefit to human, social and environmental wellbeing?	Environmental impact assessment	Does the company have targets/ strategies in place to reduce environmental impact and/or increase the positive impact over time?	Energy usage Greenhouse gas emissions
		Social impact assessment	Does the company assess the broader societal impact of the AI system's use beyond the individual user?	Change in number of employees Cost saving from AI
Human-centred values: AI systems should respect human rights, diversity, and the autonomy of individuals. AI systems should be designed to augment, complement and empower human cognitive, social and cultural skills.	Are the company's AI systems assessed to respect human rights, diversity and autonomy?	Human protection	Does the company have policies and identify requirements to protect stakeholders, particularly data subjects and individuals affected by the AI systems (decisions/outputs)?	N/A
		Human rights	Does the company embed AI within its human rights and modern slavery strategy and disclosures?	Number of AI risks (human rights) Number of audits for AI risks (human rights)
Fairness: AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups.	Has the AI system been designed and deployed to minimise bias and promote inclusion and fairness?	Diverse team	Does the company have a diverse team in place to design, develop, deploy and operate AI systems?	Diversity metrics (e.g., AI teams, diversity in AI risk committee)
		Bias	Does the company have guardrails in place to mitigate the risks of bias (e.g., racial, gender) in the datasets used for the AI system?	Diversity metrics (e.g., gender diversity, demographic diversity, geographic diversity in data set)
		Inclusion	Does the company integrate inclusion and accessibility in the design and deployment of AI projects and is this tested throughout the lifecycle?	N/A
Privacy/security: AI systems should respect and uphold privacy rights and data protection, and ensure the security of data.	How do the AI systems elevate the company's data security risk, has this been assessed and what action has been taken to mitigate this risk?	Cyber security	Does the company have proper measures to prevent and control for attacks?	Number of cybersecurity incidents related to AI systems
		Copyright protection	Does the company ensure the suitability of the data collection and the sources and document the description of data sources?	Data governance compliance rate
Reliability and safety: AI systems should reliably operate in accordance with their intended purpose.	How does the company ensure the reliability and safety of its AI system to deliver services in accordance with their intended purposes?	Quality management	Does the company have increased oversight of AI systems which are used in critical operations or assets?	Number of critical systems with AI embedded
			Does the company involve independent experts for model evaluation, particularly for a foundation model?	Independent expert rate for AI model evaluation

PRINCIPLE	PRINCIPLE QUESTION	INDICATOR	EXAMPLE SUB QUESTIONS	EXAMPLE METRICS
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.	How is the company informing its stakeholders of AI use within different arms of the business, related risks and opportunities?	Explainable system	Does the company evaluate the interpretability of the AI system if it can produce explanations about the model, data and decisions for the users?	AI decision factor (input) importance score
		User notification	Does the company inform users when they are interacting with an AI system?	Percentage of interactions where users are notified
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.	What mechanisms are in place for people to challenge the use or outcomes of the AI system to promote healthy contestability?	Internal complaints management	Does the company have complaints process in place where affected internal users can voice concerns?	Number of complaints Completion rate Time to resolve complaint
		External complaints management	Does the company have a complaints process with multiple channels in place (e.g., whistle-blower hotline, online complaint form) where affected external users can voice concerns?	
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.	Does the company have designated responsibility for AI and RAI within the organisation (person, department or committee)?	Risk management	Does the company establish methods and metrics to quantify and measure the risks associated with its AI systems?	Number of AI risk metrics (e.g., risk exposure index, risk severity score, risk monitoring frequency)
		AI incident management	Does the company have a clear reporting system or process in place for serious AI incidents to inform external stakeholders (e.g., market surveillance authorities, communities) beyond the company?	Number of AI incidents informed to external stakeholders
		Accountability framework	Does the company have an accountability framework to ensure that AI related roles and responsibilities are clearly defined?	Percentage of defined AI roles and responsibilities

As Australia's national science agency, CSIRO is solving the greatest challenges through innovative science and technology.

CSIRO. Creating a better future for everyone.

Alphinity Investment Management: Investing in quality, undervalued companies entering an earnings upgrade cycle.

Aspire. Sustain. Prosper.

Contact us

CSIRO
1300 363 400
+61 3 9545 2176
csiro.au/contact
csiro.au

Alphinity
esg@alphinity.com.au
alphinity.com.au/contact