

Human Technology
Institute

The State of AI Governance in Australia

REPORT

The Human Technology Institute (HTI) is building a future that applies human values to new technology. HTI embodies UTS's strategic vision to be a leading public university of technology, recognised for its global impact specifically in the responsible development, use and regulation of technology.

Acknowledgements

Authors: Lauren Solomon and Professor Nicholas Davis.

HTI contributors: Jemima Back, Sophie Farthing, Milan Gandhi, Professor Edward Santow, Nella Soeterboek and Llewellyn Spink.

Special thanks to the Minderoo Foundation for their generous support and funding of this program.

Thanks also to HTI partners Gilbert + Tobin, KPMG and Atlassian for their provision of expertise, legal advice, engagement with company directors and advisory support.

We are particularly grateful for the time and expertise of Verity Firth at UTS; Christian Gergis at AICD; Anna Jaffe and David Masters at Atlassian; Jessica Wyndham, Francine Hoo, Susan Staples and Georgie Withers at KPMG; Peter Waters, Simon Burns and Jen Bradley at Gilbert + Tobin; Burcu Kilic and Emma McDonald at Minderoo Foundation. Whilst they provided invaluable advice for this project, the co-authors – Lauren Solomon and Nicholas Davis – are solely responsible for the content of this report.

Citation

Solomon, L., & Davis, N., (2023) *The State of AI Governance in Australia*, Human Technology Institute, The University of Technology Sydney

© Human Technology Institute, The University of Technology Sydney.

Foreword	2
<hr/>	
1. Executive summary	4
<hr/>	
2. What is AI, and where is it being used?	8
2.1 How can we define AI for governance purposes?	10
Box 1: What kinds of systems are usefully defined as AI?	11
2.2 Why and how is AI being used by organisations today?	12
<hr/>	
3. Harms, risks, and perceptions of AI systems	14
3.1 Where do AI risks and harms come from?	15
3.2 Common harms to individuals from AI systems	16
3.3 Risks to organisations from AI systems	18
3.4 Societal risks	20
3.5 Public perceptions of AI systems	22
Box 2: Governance implications of generative AI systems	23
<hr/>	
4. How are organisations currently governing AI systems?	26
4.1 The need for AI-focused governance	27
4.2 Current approaches to AI governance	28
Box 3: The promise and shortcomings of governance via AI principles	31
<hr/>	
5. Legal obligations of Australian organisations using AI	32
5.1 Duties of company directors related to the use of AI	35
5.2 Legal obligations for Australian organisations using AI	38
<hr/>	
6. International trends in AI law and policy	50
6.1 The global trend towards AI regulation	51
6.2 Key features of international AI initiatives and interventions	53
6.3 The role of international standards in AI governance	55
<hr/>	
7. Actions for corporate leaders	56
<hr/>	
Appendix: Common principles of responsible or ethical AI	63
<hr/>	
Endnotes	64

Acknowledgement of Country

UTS acknowledges the Gadigal people of the Eora Nation, the Boorooberongal people of the Dharug Nation, the Bidiagal people and the Gamaygal people upon whose ancestral lands our university stands. We would also like to pay respect to the Elders both past and present, acknowledging them as the traditional custodians of knowledge for these lands.



Foreword

Artificial intelligence (AI) is essential to how organisations operate and make decisions today. Any corporate leader who thinks their company does not rely on AI is almost certainly unaware of how AI is already built into many of their organisation’s systems or how AI is being used without their knowledge.

Company directors and senior executives each have legal obligations regarding the use of AI in their organisations. The challenge for corporate leaders is understanding how to use AI lawfully and responsibly. This report will assist corporate leaders in rising to that challenge.

Australia’s corporate leaders play a critical role in ensuring that the AI systems used in their organisations are legally compliant, fair, fit-for-purpose, accurate and accountable. As the use of AI rapidly expands, corporate leaders are shaping – and responsible for – how AI systems are deployed.

This report forms part of the Human Technology Institute’s (HTI’s) Artificial Intelligence Corporate Governance Program (AICGP). The AICGP is helping Australian organisations capitalise on the opportunities offered by AI systems while addressing the commercial, regulatory, and reputational risks that AI systems pose.

As with all of HTI’s work, this report is focused on the human implications of corporate use of AI. When organisations understand those human implications and respond with appropriate governance measures, they can adopt AI safely and effectively.

Through this report and other AICGP workstreams, HTI is supporting corporate leaders to deepen their understanding of the AI landscape, understand current and evolving legal obligations, and identify new approaches across the corporate governance ecosystem to better serve Australians and corporate leaders’ own organisations.

More than two-thirds of Australian businesses report using or actively planning to use AI systems in their business operations. Companies are investing in AI systems – often through a mix of internal projects and external procurement – to help teams predict outcomes, optimise operations, draw conclusions from data, generate content or make decisions.

These investments promise significant benefits to the Australian economy. CSIRO’s Data61 has predicted that AI and other digital technologies will contribute an additional \$315 billion in economic activity by 2028.

However, we have also seen the disastrous consequences of failure caused by AI and algorithmic systems. Those consequences can include significant harm to individuals and groups, costly commercial outcomes, reduced shareholder value, and declining consumer and stakeholder trust. The ‘Robodebt’ scandal showed dramatically how algorithmic systems can go wrong at scale.

Robodebt was also an example of how the harm resulting from technology misuse is often experienced disproportionately by people who are least able to influence the design and use of new technology. Hence, the AICGP aims to support organisations in ensuring that harms and risks facing marginalised communities are identified, avoided and governed.

Unfortunately, Australian organisations currently lack the awareness, skills and processes to avoid future Robodebt-style failures in the private and public sectors. As Australian organisations increasingly deploy AI systems, corporate leaders must match investment in developing and procuring AI systems with appropriate governance systems.

While Australia has entered the ‘age of AI deployment’, there has been limited enforcement of laws relating to cyber security, human and consumer rights, competition, and negligence. This has led some to conclude that companies and governments operate in a ‘digital wild west’ where the law doesn’t apply to AI. This is wrong. Australian regulators and courts are increasingly ramping up enforcement of existing law in the context of AI. The regulatory risk for how companies use and rely on AI, as well as the law itself both internationally and locally, is rapidly evolving.

The AICGP is emblematic of how HTI works with partners to put humans at the centre of emerging technologies by building strategic skills, designing nuanced policies and developing practical tools. We thank the Minderoo Foundation and HTI partners, Atlassian, Gilbert + Tobin and KPMG, for their expertise and support.

Professor Nicholas Davis
Professor Edward Santow
Lauren Solomon

Part 1.

Executive summary

This report provides an overview of the current state of corporate governance concerning artificial intelligence (AI). We focus on the practices and obligations of company directors and senior executives of organisations deploying AI in Australia. Our findings are based on surveys, structured interviews, and workshops engaging over 300 Australian company directors and executives, as well as expert legal analysis and extensive desk research.

Company directors and senior executives bear distinct and interrelated legal duties and organisational responsibilities regarding AI. This report uses the term ‘corporate leaders’ to refer to both groups. It uses the term ‘organisation’ to refer to different structures, such as publicly listed and private companies, partnerships, co-operatives, trusts, and joint ventures, as well as not-for-profit organisations and public entities.

Corporate leaders across Australia are increasingly aware of the potential of AI systems to create commercial value. However, they are also increasingly cognisant of risks that can flow from AI system failure, misleading or malicious use, and overuse.

While the regulatory environment related to AI systems is evolving, when an organisation uses or relies on an AI system that causes harm, the organisation will generally be responsible. To ensure that AI systems are accurate, accountable, fair and fit-for-purpose, Australian organisations must match their growing investment in technological systems with a corresponding transformation of their governance systems.

How and why is AI being used by organisations?

AI is rapidly becoming an essential part of how Australian organisations operate. While research suggests that Australia lags behind many other developed nations in AI uptake, at least two-thirds of Australian organisations are already using, or actively planning to use, AI systems to support a wide variety of functions. Organisations are introducing AI systems to improve productivity, achieve process efficiencies, improve customer service, and create new products.

The use of AI systems within organisations is shifting in two important ways. First, AI is being bundled into products and services that organisations procure through technology partners, and being used by employees and across supply chains in ways that are often not fully visible. Second, AI systems are being applied closer to the ‘core’ of organisations, with rapid growth in the adoption of AI systems in strategy, corporate finance, and risk functions.

Although AI systems are central to how organisations operate, most corporate leaders across Australia are unaware of where and how AI is being used to create value. In addition, corporate leaders across Australia report that they lack the awareness, skills, knowledge, and frameworks to guide responsible AI investment and use effectively.

Australia’s corporate leaders play a critical role in ensuring that the AI systems used in their organisations are legally compliant, fair, fit-for-purpose, accurate and accountable.



What harms and risks arise from AI systems?

While the opportunities associated with AI are real, so are the risks and harms. AI-related risks and harms flow from three sources: AI system failures, the malicious or misleading use of AI systems, and the overuse or reckless use of AI systems.

As the deployment of AI systems accelerates, organisations are increasingly exposed to AI-driven commercial, regulatory, and reputational risks. Meanwhile, individuals and communities can and do suffer irreversible harm. At a societal level, poorly-governed AI systems can amplify inequality, undermine democracy, contribute to unemployment, threaten Australia's security and increase social isolation.

Australians are increasingly concerned about AI-related risks. Only a third of Australians say that they trust AI systems, and less than half believe the benefits of AI outweigh the risks.

How are organisations currently governing AI?

AI-related harms are not inevitable nor unforeseeable. Unfortunately, current organisational risk management and governance approaches are inadequate to address those harms.

Australian organisations generally lack a systemic governance or risk-management approach to identify and address AI-related harms and risks. A significant proportion of AI-related use – including systems embedded within suppliers and used without authorisation by employees – is not recognised or captured by current governance processes. Corporate leaders report that they lack the awareness, skills, knowledge, and frameworks to guide responsible AI investment and use.

A common form of AI-specific governance is the adoption of a set of ethical or responsible AI principles. However, evidence suggests that principles are necessary but insufficient. Executives and teams across organisations report that such principles alone do not help them make practical decisions about procuring, designing, deploying, and managing AI systems.

What obligations apply to corporate leaders and organisations using AI in Australia today?

Given the growing prevalence of AI systems, corporate leaders must understand the current and evolving rules governing their use in Australia and other important markets.

Australia currently has very few laws that are directed expressly towards AI systems. This has led some corporate leaders to wrongly assume that AI systems are generally unregulated or that AI use is primarily a question of ethics. In fact, companies' development and use of AI are regulated in Australia by a range of technology-neutral laws of general application. Moreover, regulators are increasingly enforcing these laws more effectively. As a result, companies should anticipate a more rigorous application of existing laws to AI systems.

Under section 180 of the *Corporations Act 2001* (Cth), company directors have a personal legal duty to ensure that effective risk management and compliance systems are in place. It is increasingly clear that company directors must exercise reasonable care and diligence to ensure there are appropriate oversight and governance systems for the AI systems their companies rely on. As companies rely on AI more and more in their operations, company directors should be alert to the rising need to modernise and adapt their companies' oversight and governance systems to account for change to their operations.

More generally, organisations bear legal obligations that pertain to AI systems. The failure, malicious use or overuse of AI systems may breach laws related to privacy, consumer protection, anti-discrimination, negligence, cyber security, and work health and safety, as well as a range of industry-specific regulations in sectors such as finance and healthcare.

What should corporate leaders expect from AI regulation?

Leading jurisdictions are rapidly enacting laws and policies that encourage positive innovation in AI while protecting people from harm. Some reforms seek to modernise and clarify technology-neutral rules to account for the unique characteristics of AI systems. Other reforms aim to regulate AI directly.

Australian governments have been comparatively slow to consider reform that responds to the rise of AI. However, Australian policy makers are exploring AI-specific laws in line with international principles for how best to regulate AI systems. This includes discussion around risk-based regulation, sector-specific requirements, use case restrictions, and new proposals to reform laws on privacy, intellectual property, liability, cyber security, and transparency.

Furthermore, an evolving set of international standards are setting baseline expectations in markets regarding how companies should approach the governance and risk management of AI systems.

What actions should corporate leaders take to govern AI effectively?

To govern AI appropriately, corporate leaders need to invest in four areas:

1. capacity building and developing strategic expertise related to AI
2. creating a suitable AI strategy
3. implementing governance systems that effectively address risks associated with AI
4. supporting a human-centred culture regarding their use of AI.

First, corporate leaders should invest in 'strategic expertise' concerning AI across the organisation. While many organisations have invested heavily in acquiring technical data science skills and capabilities, there is a critical shortage of strategic AI knowledge and experience among non-technical teams involved in decision-making or use of AI systems. Given how essential AI systems are to organisations today, corporate leaders, operational teams and front-line staff need a 'minimum viable understanding' of how AI systems work.

Second, corporate leaders should ensure that their organisation has a comprehensive AI strategy that prioritises opportunities, uncovers potential harms and risks, recognises legal obligations, and establishes a risk appetite for AI deployment. The strategy should be aligned to broader organisational objectives, as well as existing policy frameworks and risk and assurance practices. This strategy should be a dynamic document, able to be updated as novel AI approaches become available, risks and opportunities emerge or organisational risk appetite changes.

Third, corporate leaders should design and implement an integrated, structured, and comprehensive governance system for AI systems. Such a governance system should, at a minimum, establish clear and accessible processes, policies and standards, including mechanisms for oversight and assurance, document systems, identify potential impacts, determine legal requirements, and establish appropriate delegations and accountability for failures, malicious use and overuse.

Fourth, corporate leaders should support the development of a human-centred culture regarding the development and use of AI. AI systems should deliver value to all stakeholders, including employees. Staff members and customers should feel that AI systems serve their interests rather than the inverse.

Shaping the future of AI governance

Corporate leaders should recognise that thoughtful and effective AI regulation will be central to Australia's ability to benefit sustainably from AI systems. Corporate leaders need to be rigorous in understanding their legal and ethical obligations. They should be proactive in engaging with their applicable peak or professional bodies, as well as relevant regulators, communities, and other stakeholders, to ensure they are effectively discharging their obligations.

Part 2.

What is AI and where is it being used?

Australian organisations are rapidly deploying AI systems across all sectors for an expanding set of purposes.

This investment in AI is forecast to raise productivity and expand economic output. The economic advantages of increased productivity driven by AI are projected to contribute an estimated \$6.6 trillion to the global economy by 2030.¹

There is emerging evidence that investing in AI systems adds significant value to organisations. A 2022 survey by GitHub found that 88% of programmers feel more productive when using a generative AI system that supports code generation and completion.² Meanwhile, a 2023 study on the use of generative AI to help customer support

agents found that access to AI assistance increases worker productivity by 14 percent, enabling them to resolve more customer issues per hour.³ Around the world, 70% of companies adopting AI in marketing, sales or product development report revenue increases, while 30% see cost reductions.⁴

Realising the promise of AI through effective governance requires a shared understanding of three distinct aspects of AI: how AI is defined, the current and likely use cases of AI systems, and the individual harms and collective risks that can arise from AI use.

A 2023 study on the use of generative AI to help customer support agents found that access to AI assistance increases worker productivity by 14 percent.



2.1 How can we define AI for governance purposes?

There is no standard or universally agreed definition of ‘artificial intelligence’.⁵

This is partly because AI is a vast field encompassing a wide range of techniques.⁶ Organisations employ systems that rely on very different AI approaches. These include logic-based, symbolic systems (sometimes known as ‘Good Old-Fashioned AI’), probabilistic models such as Bayesian machine learning, and so-called ‘connectivist’ approaches like deep learning.

The phrase ‘artificial intelligence’ itself is also responsible for generating confusion. When examined closely, AI systems are neither artificial nor intelligent.⁷ Furthermore, our sense of what is evidence of intelligence when applied to digital systems is constantly shifting.⁸

Nevertheless, building effective governance systems requires corporate leaders and teams across an organisation to broadly agree on what is considered an AI system.

A helpful definition, adapted from work by the EU and OECD, is the following:⁹

*Artificial intelligence (‘AI’) is a collective term for machine-based or digital systems that use machine or human-provided inputs to perform advanced tasks for a human-defined objective, such as **producing predictions, advice, inferences, decisions, or generating content.***

Some AI systems operate autonomously and can use machine learning to improve and learn from new data continuously. Other AI systems are designed to be subject to a ‘human in the loop’ who can approve or override the system’s outputs. AI systems can be custom developed for a specific organisational purpose. Many are embedded in products or deployed by suppliers in upstream or outsourced services.

As AI advances rapidly, corporate leaders would be well-served to take a broad view of what constitutes an AI system within their organisation.¹⁰ Box 1 provides a non-exhaustive list of systems that meet the definition of AI above.

AI systems differ in important ways from traditional IT systems. Most rely heavily on models that are a result of training data processed by algorithm, rather than logic-based programming. Operating them tends to require larger amounts of data, some of which may be live or unstructured, and is costly to manage and secure. AI systems also tend to be less transparent than traditional software systems, more challenging to test and pose greater difficulties in predicting failures. They often require more frequent maintenance and oversight.¹¹

Recognising that organisationally relevant AI systems may operate outside immediate corporate boundaries is also essential. AI systems are often embedded in outsourced or third-party-provided functions on which organisations rely. For example, recruiting partners, advertising platforms and translation services can employ AI systems to make critical decisions for which an organisation – and its officers – are responsible.

Box 1: What kinds of systems are usefully defined as AI?

- **Machine learning systems** – a broad set of models that have been trained on pre-existing data to produce useful outputs on new data.
- **Expert systems** – systems that use a knowledge base, inference engine and logic to mimic how humans make decisions.
- **Natural language systems** – models that can understand and use natural language and speech for tasks such as summarisation, translation, or content moderation.
- **Facial recognition technologies** – systems that verify a person, identify someone, or analyse personal characteristics using face data drawn from photos or video.
- **Recommender systems** – systems that suggest products, services or information to a user based on user preferences, characteristics, or behaviour.
- **Automated decision-making systems** – systems that use data to classify, analyse and make decisions that affect people with little or no human intervention.
- **Robotic process automation** – systems that imitate human actions to automate routine tasks through existing digital interfaces.
- **Virtual agents and chatbots** – digital systems that engage with customers or employees via text or speech.
- **Generative AI** – systems that produce code, text, music, or images based on text or other inputs.
- **AI-powered robotics** – physical systems that use computer vision and machine learning models to move and execute tasks in dynamic environments.

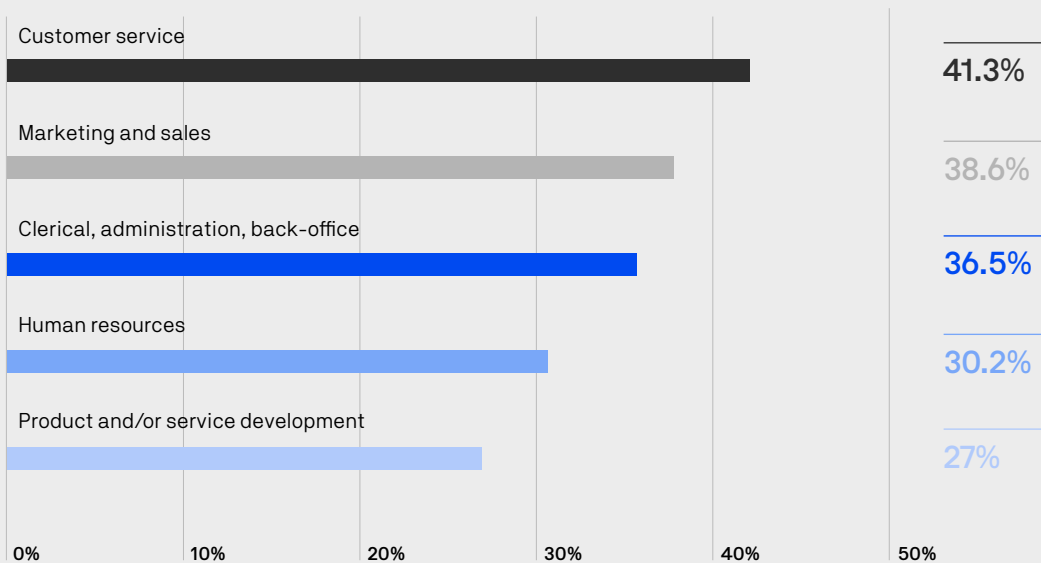
2.2 Why and how is AI being used by organisations today?

AI is rapidly becoming core to how organisations in Australia and around the world operate. Research conducted by HTI from December 2022 to April 2023 found that nearly two-thirds (64%) of Australian corporate leaders and strategic decision makers surveyed were either already using or planning to use AI systems in their operations.¹² 37% of respondents reported that their organisation currently used AI in limited ways, while 8% used it in multiple ways across functions and teams. This finding is broadly consistent with trends from other similar surveys, which also reveal high levels of variation across industries.¹³

AI is being adopted because it promises to add significant value to organisations. The three most common reasons for Australian organisations to adopt AI are process efficiency, increased productivity, and better customer experience. HTI research suggests that senior executives tend to favour process efficiencies as a dominant driver of AI investment, while directors place significantly more weight on the potential for AI systems to contribute to better customer experience and productivity gains.

In line with other surveys, HTI research found that the top five operational areas prioritised for AI deployment by Australian business leaders were customer service, marketing and sales, clerical and administrative, human resources, and product and/or service development (Figure 1). This tracks closely to global trends.¹⁴

Figure 1 – Top five use cases for AI in Australian organisations (HTI, 2023)



Percentage of responses to survey question: 'In which operational areas are you using or planning to use AI in your organisation?'

Australia currently lags many other jurisdictions, such as France, Germany, the US, Israel, the UK, and Canada, in AI investment and use.¹⁵ However, HTI interviews and survey data suggest this is changing. The public release of generative AI tools is contributing to rising levels of interest across corporate leaders. Furthermore, HTI research suggests that the use of AI systems in enterprises is broader than survey data suggests.

Global and Australian data on AI adoption indicate that this widening use of AI systems is due to two crucial shifts taking place in the corporate use of AI. These present an urgent need for organisations to invest in structured, fit-for-purpose AI governance.

First, AI has entered the 'era of deployment'. Until relatively recently, investing in enterprise AI systems required extensive teams of data scientists, engineers, and developers to acquire and clean data, create or train custom AI models from scratch, and build supporting interfaces. The advent of cloud-based machine learning operations (MLOps) and data management platforms enables teams to access, develop, operate, and manage AI models and associated data with far greater flexibility. AI systems are increasingly provided via third party products and services, particularly in recruitment, customer relationship management, cyber security, and customer service systems.

Second, AI adoption is transforming the core of business operations. While customer service operations, marketing and sales and product development remain the most popular use cases for AI worldwide, the rate of AI adoption is growing most rapidly in human resources, risk, strategy and corporate finance.

An important finding of HTI's research that further support an underestimation of AI use is that corporate leaders – even technically-focused executives such as chief technical officers and chief data officers – report being unsure about the scope and scale of their organisations' AI use. This suggests that many surveys will tend to underestimate AI use across organisations.

In HTI interviews, corporate leaders report that siloed decision-making across divisions and a lack of consistent reporting prevent them from gaining a holistic view of AI use. Third-party services that employ AI are regularly overlooked.¹⁷ Moreover, the recent popularity of generative AI systems such as OpenAI's ChatGPT has led to an explosion in 'shadow AI': employees' unauthorised use of cloud-based AI applications for work-related purposes.¹⁸

AI use in Australian organisations can be characterised as rapidly increasing, systematically under-recognised, increasingly embedded and closer to the core of how value is created.

AI use in Australian organisations can be characterised as rapidly increasing, systematically under-recognised, increasingly embedded and closer to the core of how value is created.

Part 3.

Harms, risks, and perceptions of AI systems

Without proper governance, the rapid deployment of AI systems exposes organisations, employees, consumers, and the broader community to severe harms and significant risks. Organisations responsible for AI-related harms will be exposed to commercial, regulatory, and reputational risks. As the Robodebt scandal showed, the failure of even relatively unsophisticated systems can result in catastrophic consequences at scale.

In this report, we deliberately distinguish between the harms AI systems can impose on individuals and groups and the risks that can emerge at organisational and societal levels.¹⁹ Harms are typically real, concrete and ‘vested’. People experience harm and bear lasting consequences as a result. By contrast, risks tend to be future-oriented, distributed and characterised by uncertainty.²⁰

This distinction is essential for effective governance. In the same way that health and safety regulation is designed to prevent harm to individuals, corporate leaders should recognise that AI systems can and do harm individuals and groups.

Corporate leaders should also acknowledge that these harms disproportionately affect marginalised groups. Not only do vulnerable or marginalised individuals possess fewer resources to absorb the cost of AI system failures or to seek redress, their lack of voice and power makes it less likely that their stories of harm will be heard at all. Furthermore, marginalised groups are also more likely to be subject to the overuse of AI and less equipped to assert their rights in response.²¹

To discharge their legal duties, organisations developing or deploying AI systems (and their officers) must therefore carefully and systematically invest in identifying and mitigating the individual harms, organisational and societal threats posed by AI systems.

3.1 Where do AI risks and harms come from?

AI systems create or exacerbate harms and risks through three primary pathways, as shown in Figure 2. Each of these pathways can create direct harms or limit human rights for individuals. These pathways also create risks for organisations and societies more broadly.

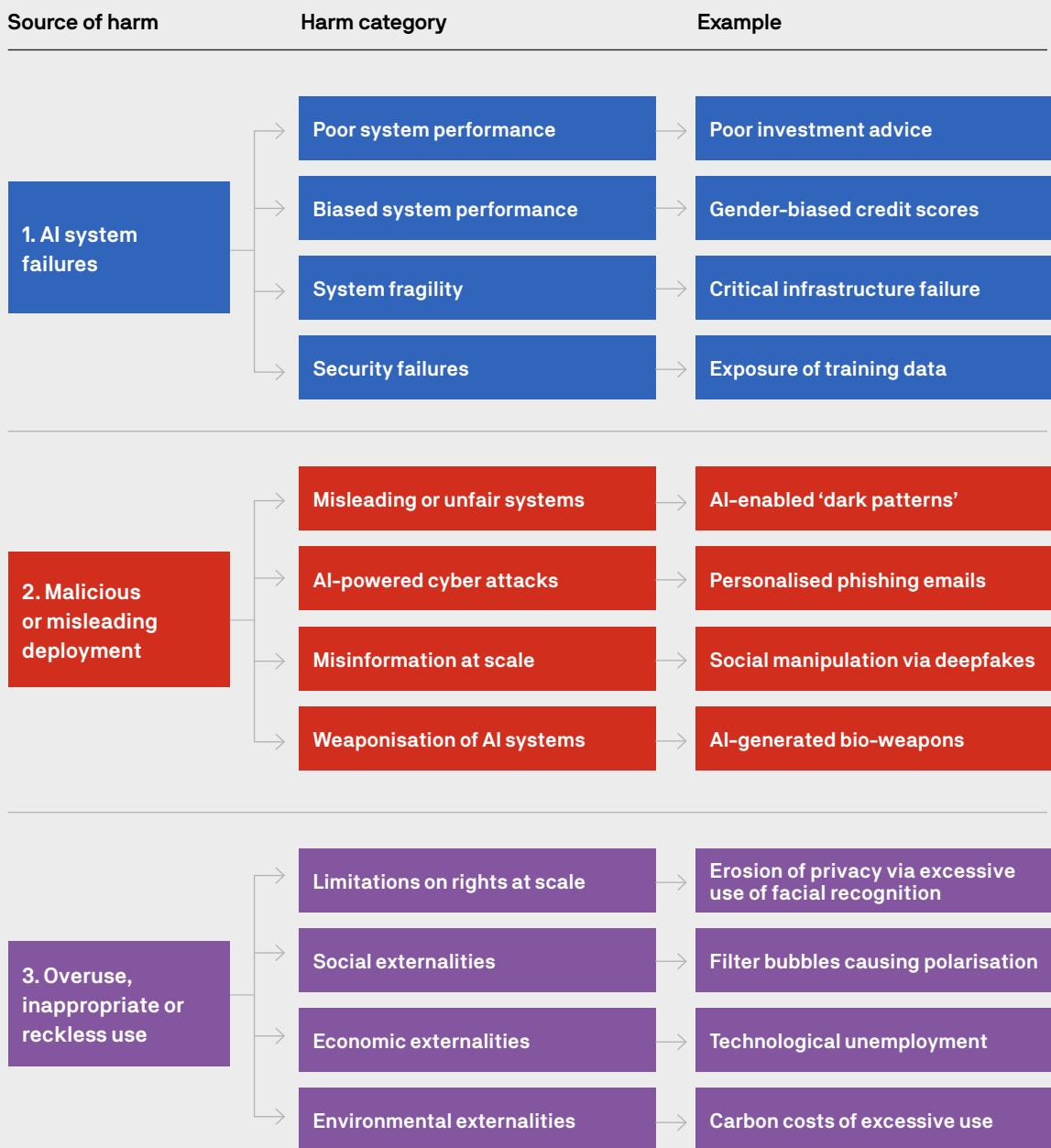
First, AI systems cause harm and risks when they fail to operate in the way, or to the level of quality, required. These failures include when overall system performance is sub-optimal, when errors are systemically distributed in ways that create bias or result in discrimination, when the system is fragile, or when the system is insecure.

Second, AI systems cause harm when they are deployed for malicious purposes or in misleading ways. This category includes when AI systems are designed to mislead users, the use of AI systems to perpetrate cyber attacks or generate misinformation, and when AI systems are used to create outputs directly intended to harm people, such as bio-weapons.

Third, AI systems cause harm when they are over-used, used inappropriately or deployed recklessly without regard to their second- and third-order effects. This category includes when the excessive use of AI technologies such as facial recognition or predictive policing at scale severely limits human rights. In this category are also so-called ‘unintended consequences’ – social, political, economic, and environmental impacts of AI systems that developers, deployers or users fail to account for or recognise.

Corporate leaders should acknowledge that AI system harms disproportionately affect marginalised groups.

Figure 2 – Pathways of harms and risks flowing from AI systems



3.2 Common harms to individuals from AI systems

Organisations regularly deploy AI systems in ways that directly impact the experiences of consumers and employees. System failures, malicious deployment, or overuse can cause direct harm to individuals.

Corporate leaders should be cognisant of these harms. In addition to any ethical preference to avoid harming people, harms created by AI systems can create commercial, reputational, and regulatory risks for the organisation. Table 1 highlights some key potential harms to individuals that can arise from organisations using AI systems.

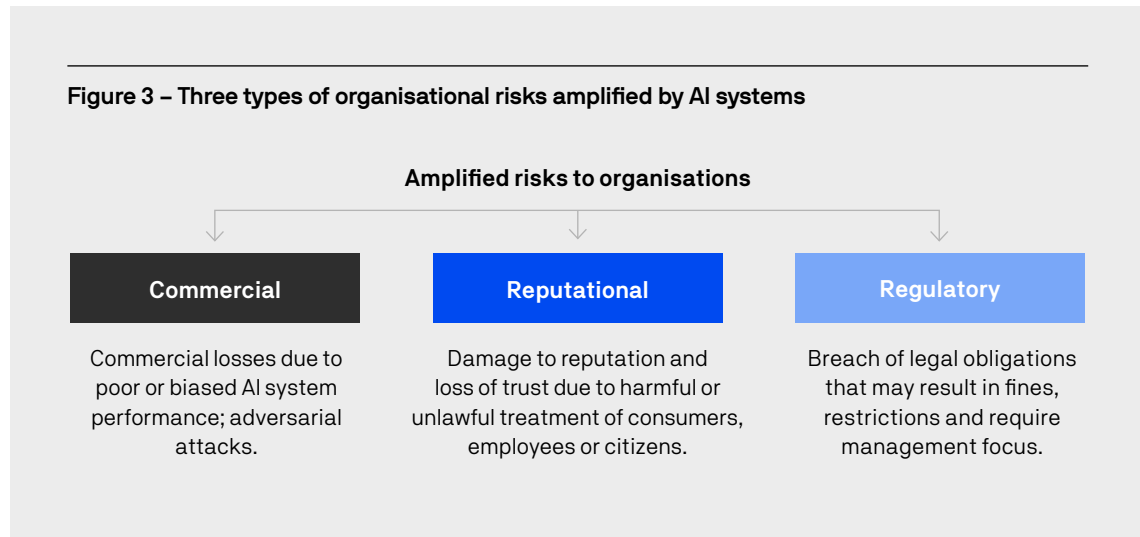
Table 1 – Selected individual harms from AI systems

Harm category	How harms to individuals may arise	Examples
Poor system performance	A user or affected individual can be physically harmed, have their property damaged by a system, experience psychological distress or suffer reputational harm because of errors in an AI system's output.	In 2018 a self-driving vehicle failed to detect a female pedestrian, hitting and killing her. ²² In 2023, ChatGPT hallucinated a sexual harassment allegation against a Law Professor at George Washington University, causing him distress. ²³
Biased system performance	AI systems can fail in ways that result in individuals being unfairly denied access to a service or experiencing systematically worse treatment based on a protected attribute. Errors in an AI system can also result in an individual's liberty or bodily autonomy being unjustifiably restricted or violated, e.g. false arrests, or the use of biased algorithms in bail or sentencing decisions.	A facial recognition technology (FRT) system in the US misidentified a Black teen, barring her from accessing an ice-skating rink. ²⁴ Research shows that algorithms used in FRT are significantly less accurate when distinguishing between people of colour, women, and children. ²⁵ This could result in breaches of anti-discrimination laws.
Security failures	An individual's valuable or essential data can be accessed, extracted, altered, or restricted by unauthorised third parties thanks to the poor design or operation of an AI system.	Third parties routinely seek to hack or compromise the integrity of the AI system's decision-making process. ²⁶
Misleading systems	Artificial intelligence systems can – either deliberately or inadvertently – provide information or advice to consumers which is misleading or deceptive, e.g. product rankings and recommendations which are incorrect. Harm arises when this information is relied upon to the customers' detriment, for example by causing them to pay more. An AI system may also be designed to unfairly result in harms such as poorer service, higher costs, or the inability to exercise choice or consumer rights.	Trivago breached s 18 of the Australian Consumer Law (ACL) and engaged in misleading or deceptive conduct because 66% of the time its algorithm would nominate a hotel offer more expensive than the best offer. ²⁷ 83% of Australians have experienced negative consequences from digital design features designed to influence their behaviour. Organisations that deploy AI-enabled 'dark patterns' may be in breach of ACL. ²⁸
AI-powered cyber attacks	AI systems can be used to target vulnerable individuals at scale. The use of highly personalised content and 'deepfake' audio or video can be used to commit cybercrimes or intimidate a natural person.	AI is being used to develop targeted phishing emails that are more likely to be opened than human-generated ones. ²⁹
Reckless or otherwise unlawful limitations on rights	Some AI systems – such as facial recognition technologies – can harm individuals by breaching their rights when used without the full, prior, and informed consent of affected individuals.	Following reports by CHOICE, in 2022, the Office of the Australian Information Commissioner (OAIC) opened investigations into retailers Bunnings and Kmart Australia, focusing on the companies' use of facial recognition technology. ³⁰

3.3 Risks to organisations from AI systems

Types of organisational risks

In general, AI systems expose organisations to three types of risks: commercial, regulatory, and reputational (Figure 3).



- **Commercial risks** occur when AI systems directly lead to sub-optimal decisions, products, or services. Up to 85% of AI projects fail to live up to their promise,³¹ meaning that developing or procuring an AI system requires careful consideration of benefits and costs. Systems that perform poorly can lead to missed or lost commercial opportunities and additional costs in the form of workarounds. Furthermore, the additional complexity of AI systems compared to other digital systems can cause organisations to generate a form of ‘technical debt’ through significant and ongoing integration challenges, maintenance, and upgrade costs.
- **Reputational risks** occur when system failures, malicious use, or overuse create harms experienced by or visible to external stakeholders. The most common example is when AI systems breach customers’ expectations around their privacy.³² Similarly, perceptions of algorithmic bias can lead to reputational damage, as when the Apple Credit Card seemed to provide larger credit lines to men than women.³³ An organisation may also suffer reputational damage when it cannot adequately explain its AI system’s decisions.³⁴
- **Regulatory risks** occur when an organisation’s use of AI – or the harms to individuals that may result – breaches its legal obligations. These are discussed in detail in section 5 below.

These three forms of organisational risk often co-occur. For example, AI systems that unfairly discriminate against job applicants risk breaching anti-discrimination laws. They may also undermine the commercial and cultural benefits of a diverse workforce and cause an organisation to miss out on hiring better-qualified candidates. A fault in a self-driving vehicle that results in injury may give rise to a wide range of commercial, reputational, and regulatory risks, all of which may undermine the trust and confidence of consumers. Organisations using AI systems to generate content for their business using generative AI may produce poor quality output, be exposed to intellectual property risks, and suffer reputational damage for failing to disclose their use of machine-generated text or images.³⁵

The inability to explain how an AI system generated an unfavourable result for customers or employees may contribute to a decline in customer satisfaction, perceptions of unfairness and accusations of misconduct. Hence, the intrinsic opacity of some AI systems may lead to reputational damage and legal action. This risk can be heightened by a failure to adequately train employees in how the AI system works, where human decision-makers show too much deference to AI system recommendations, or by the selection of AI models which lack explainability as a key feature. This is of particular concern in circumstances where being able to provide reasons for decisions is a legal requirement.³⁶

Organisational awareness of AI-related risks

HTI research indicates that corporate leaders' perceptions of organisational risk change with the level of experience with AI use. HTI's survey of corporate leaders found that, while those who reported limited use of AI within their organisation mostly perceived AI risks as 'low – moderate', those who reported *using AI in multiple ways* across their organisation were far more likely to characterise perceived risks as either 'very high' or 'very low' (Figure 4). For inexperienced AI users, we see a normal distribution of perceived risks. For experienced corporate leaders, the distribution becomes bimodal.

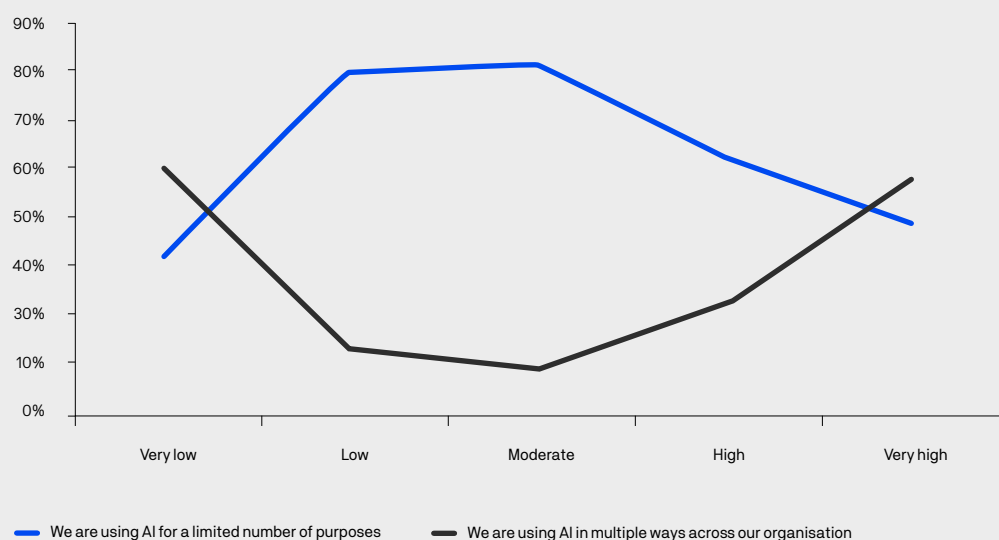
This finding suggests an interesting 'experience effect' with regard to risk perception. It suggests that corporate leaders build a more nuanced understanding of the potential significant scale and scope of risks posed to their organisation with increased exposure and use. It may also suggest overconfidence on the part of some respondents in relation to familiar AI systems that have been deployed without incident to date. Finally, the

finding highlights the challenge of prioritising the implementation of AI risk management systems when organisations are in the nascent stages of AI deployment.

These data on risk perception must be interpreted in light of the fact that corporate leaders tend to report low levels of expertise in the field of AI and low awareness of the AI systems being deployed by their organisation.

Organisational risk awareness may be further compromised by the fact that a rising proportion of AI systems are outside of the organisation's direct control. AI systems are frequently embedded in tools or applications acquired from or used by third parties across the supply chain. Corporate leaders and organisations more broadly should consider their awareness of and responsibilities in relation to third-party provided or deployed AI systems. This includes appreciating the various responsibilities borne by system developers, deployers and users across the AI lifecycle, as well as the rights of affected individuals.

Figure 4 – Perceptions of organisational AI risk change with experience



Strategic options and risk tolerance

Corporate leaders should recognise that there are risks associated with deciding not to use AI systems. Companies will naturally seek to adopt new technologies, including AI systems, where doing so lawfully and responsibly will add value by expanding market access, increasing efficiency, boosting productivity, or delighting customers. A failure to develop and use AI may therefore be a strategic risk for the organisation.

Corporate leaders therefore need to ensure that the board is well informed about the harms, risks and opportunities associated with the organisation's current and future use of AI systems – including the risks associated with failing to adopt AI where it is warranted. Decisions around the adoption of AI systems should be guided by an AI strategy that, inter alia, clearly defines the level and quality of risk the organisation is willing to accept in pursuit of its objectives.

3.4 Societal risks

AI can be a powerful tool for social good. AI systems have been shown to assist pandemic preparedness and response,³⁸ improve accessibility for people with disabilities,³⁹ and support conservation efforts.⁴⁰

However, corporate leaders should be mindful of the societal risks that can emerge from AI system use at scale (Table 2).⁴¹ These risks may materialise because of AI system failures, due to malicious or misleading deployment, or through overuse, inappropriate or reckless use.

While many examples of systemic bias in algorithms are associated with automated decision making in the public sector, societal risks are not limited to government or law enforcement use of AI systems. They can occur at comparable scales, 'with potentially less justification and access to fewer available remedies', when used by the private sector.⁴²

A failure to attend to societal risks may directly impact an organisation's commercial interests. For example, an organisation's use of carbon-intensive AI systems may affect its value when assessed against an ESG framework.⁴³ Societal risks that emerge in the form of widespread harms across groups will inevitably impact the organisation's reputation and create significant regulatory risks.



Societal risks may materialise because of AI system failures, due to malicious or misleading deployment, or through overuse, inappropriate or reckless use.

Table 2 – Selected societal risks from AI systems

Risk category	How risk arises	Examples
Biased system performance	Systematic bias can be embedded in AI systems in many forms beyond the existence of unrepresentative data. This includes inflection points such as how investors perceive the problem to be solved, the design of the algorithm, where developers choose to pilot a solution, and how deployers roll out systems at scale. ⁴⁴	An algorithm used on more than 200 million people in US hospitals to predict which patients should receive additional medical care was found to favour white patients over Black patients twice as often as should be the case. Bias occurred because the algorithm used health costs as a proxy for health needs, thereby incorporating real-world bias. ⁴⁵
Misinformation at scale	AI systems can be used maliciously at scale by groups seeking to self-interestedly spread propaganda and influence public opinion in ways that can spread harmful information and decrease trust. ⁴⁶	AI-generated videos have been used to spread disinformation in the Russia-Ukraine war, including deepfake videos of President Vladimir Putin declaring peace and President Volodymyr Zelenskyy telling Ukrainian citizens to surrender. ⁴⁷
Reckless or otherwise unlawful limitations on rights	AI systems can limit individual rights and liberties, undermine elections and trust in institutions, intensify tension between social groups, and erode the rule of law. Such effects will have an outsized impact on groups fighting for their rights to be recognised.	The use by governments of facial recognition systems and other surveillance technologies will in many cases undermine the right to privacy and limit freedoms of expression and assembly in ways that further diminish the voice of vulnerable groups. ⁴⁸
Social and political externalities	AI systems may contribute to and aggravate existing socio-economic inequalities by systematically disadvantaging vulnerable or marginalised populations. ⁴⁹ When used to make decisions with legal or similarly significant effect, AI systems that are not explainable or transparent may decrease accountability and create the perception of arbitrary decision-making. ⁵⁰	Due to existing over-policing trends, people of colour and low-income communities tend to be overrepresented in historical data that is used to train predictive policing algorithms. This can cause the algorithm to erroneously or unfairly identify these communities as high risk. ⁵¹ The use of AI systems in armed conflicts could create an ‘accountability gap’ in international humanitarian law, putting civilian populations at risk. ⁵²
Environmental externalities	Training and using AI systems tends to be computationally, energy and water intense. Computing demand for AI model training and inference at Meta are increasing by 150% ⁵³ and 105% ⁵⁴ respectively, while Google reports that AI systems represent 70–80% of their overall computing demand, and 10–15% of energy. ⁵⁵	Training a 2022-era generative AI model using carbon-intensive energy sources emits an estimated 500 metric tons of carbon ⁵⁶ and directly consumes 700,000 L of fresh water. ⁵⁷

3.5 Public perceptions of AI systems and governance

Australians are aware of and concerned about the harms and risks posed by AI systems.

According to recent research, only one-third of Australians say that they trust AI systems, a figure that has remained stable since 2020⁵⁹ despite AI investment growing across the economy at an estimated 20% each year.⁶⁰ Only 44% of Australians believe the benefits of AI outweigh the risks.

These levels of trust and approval vary by AI use case, with AI in healthcare trusted significantly more than AI in human resources.⁶¹ Such perceptions of risk and levels of trust are important for organisations intending to develop or procure AI systems that may impact their stakeholders.

HTI's research on facial recognition technology revealed that public perceptions of risk and corresponding levels of trust are heavily influenced by the degree to which people are aware of, are knowledgeable about and have direct experience with AI systems. When individuals participate in a simulated AI experience, and/or when the specific technology, purpose and context of a use case is explained in detail, rich and nuanced information emerges that differs substantially from broad survey data. This is particularly useful for organisations and policy makers seeking to understand why, when and to what extent a technology is perceived as low, moderate, or high risk.⁶²

Australians are demanding better enforcement of laws to respond to the risk of AI systems. Only 35% of Australians believe that current regulations, laws, and rules are sufficient to make AI use safe.⁶³ This is in line with global sentiment: around the world, 71% of people expect AI to be regulated.

In general, survey respondents do not trust developers or deployers of AI systems to govern themselves. Globally, only 26% of global respondents feel that commercial organisations can be trusted to develop, use, and govern AI in the interests of the public. When Australians were asked who should regulate AI, they preferred government, existing regulators, or an independent AI body, over the prospect of industry-led regulation.⁶⁴

Increasing the trustworthiness of AI systems will require ongoing investment in governance systems. 80% of respondents say that their trust would rise if the accuracy and reliability of an AI system were monitored. 68% said that an independent AI ethics certification would do the same, an interesting result given that such certifications still need to be fully developed and accepted across industries.⁶⁵

These findings underscore the importance of organisations developing fit-for-purpose assurance and governance mechanisms in line with their deployment of AI systems. They also provide a compelling motive for corporate leaders to engage closely with stakeholders and invest in ensuring their AI systems are worthy of public trust.

Only 44% of Australians believe the benefits of AI outweigh the risks.

Box 2: Governance implications of generative AI systems

Generative AI, a term that includes large language models (LLMs) and diffusion-based image generators, refers to a set of highly capable and flexible deep learning-based AI models and applications. Generative AI applications can generate fluent text, computer code, detailed images, convincing videos, and original music from user-provided text prompts.⁶⁶

Generative AI applications are distinct from other forms of AI. First, they are deliberately designed to generate new data based on input from the user. While traditional AI applications tend to classify, optimise, or predict primarily in order to analyse data, generative AI applications are designed to produce entirely new data in the form requested. Second, training them to be useful requires significant amounts of computational power and huge data sets.

Five characteristics of generative AI applications make them extremely useful yet fundamentally fallible and, therefore, worthy of careful study by corporate leaders and policy makers alike. These characteristics mean that generative AI systems are particularly prone to the risks of AI system failure, malicious or misleading deployment, and overuse displayed in Figure 2.

Data-related concerns: privacy, intellectual property and bias

Generative AI applications tend to be trained on large amounts of publicly-accessible data in the form of written text (including code), images, videos and music gathered from across the open internet.

Despite efforts by some developers to curate and limit training data to open-source or authorised data sets, a meaningful proportion of this data is likely to be subject to privacy law and intellectual property protections. The fact that such models have been trained on data captured without the consent of data owners may make this 'training' element of the models inherently unlawful.⁶⁷ Private data may be reproduced, and the system may reinforce underlying biases.

Generative AI applications also pose cyber security, confidentiality and intellectual property risks. Engaging with a generative AI application involves exchanging information between the user and the application. The more information that the user shares with the application, the more valuable the application's outputs are likely to be, increasing the risk that users will share legally-protected information with the application. Conversely, AI-generated works may not be eligible for legal protections, such as copyright.

In addition, the presence of stereotyped and harmful data in training sets means that the applications can be deliberately or inadvertently induced to produce biased, illegal, or inappropriate content.

Ease of use and scalability

The most recent generative AI systems are explicitly designed to be both easy to use and scalable. Many LLMs have been specifically trained to respond to conversational text inputs in any one of hundreds of languages, making the production of text, code, images, or other media extremely accessible to individuals with little to no technical training. Furthermore, many providers offer access to their systems via application programming interfaces (APIs), allowing third-party programs to access models and embed outputs into other systems quickly and directly.

This ease of use and accessibility means that generative AI systems can swiftly enter workflows and raise productivity for teams undertaking appropriate tasks, such as writing code. But it also means that, should safeguards be circumvented, generative AI systems can be used for malicious purposes by non-technical users at scale, significantly expanding the rate at which misinformation can be disseminated.

Box 2 cont.

Flexibility and use case

Generative AI systems are flexible. Trained across massive databases, they can be used to infer and generate content that is useful in a wide variety of contexts.

This versatility means that generative AI applications can be applied in an infinite array of potential use cases across sectors and organisations where text or other media is a valuable output. However, it also means that, as with other AI systems such as facial recognition, the harms and risks posed by generative AI systems are highly contextual. For example, the same generative AI application could be used to help a student learn a new programming system or to generate and distribute harmful misinformation. Governing such systems therefore requires organisations and regulators to be attuned to the use case of generative AI applications, rather than the application or model itself. At the same time, the cost and effort of using techniques such as reinforcement learning by human feedback to filter content means that developers are best placed to design ‘top down’ protections that anticipate and mitigate against a range of harmful uses.

Fluency at the expense of accuracy

The capabilities of generative AI systems, while impressive, have been optimised for fluency rather than accuracy.⁶⁸ A hallmark of LLMs is that systems can ‘hallucinate’ outputs (for example, a reference to a ‘ground-breaking’ scientific paper) that purport to be real yet do not exist or make damaging factual errors when analysing data or answering questions.⁶⁹

Even when such shortcomings – and the inherently probabilistic nature of generative AI systems – are appreciated by users, the fluency of systems can result in humans mistaking generative AI output for meaningful text, images, or audio. This will exacerbate the challenge of automation bias, where humans are inclined to trust and rely on computer-provided outputs, against their own judgement and to the detriment of stakeholders when outputs are incorrect, particularly in the absence of accountability mechanisms.⁷⁰ Should the outputs of generative AI systems be given more credence than is warranted, individuals and entire organisations may make avoidable errors, creating both harms and risks.

In the case of LLMs, Emily M Bender and others have argued that their bias toward fluency will tend to reflect and amplify biases, abusive language, or malicious content. In turn, this may produce ever-larger amounts of text that will itself be sampled in future training data: a self-reinforcing of stereotypes, problematic associations, and deceptive material.⁷¹

Emergent properties of generative AI models

Generative AI models exhibit emergence. Some abilities of generative AI applications are missing in smaller models (which are cheaper to train), yet emerge suddenly – and surprisingly – for larger models that require many millions of dollars of computing time.⁷² Emergent abilities include dramatic increases in performance, the ability to ‘reason’ more effectively, and the ability to respond to more nuanced prompts. While it is possible that a combination of open-source models and new approaches to training could alter this dynamic, the ability to access large amounts of data and computational power may remain a strategic advantage for organisations wanting to train their own generative AI models.

Box 2 cont.

Governance and regulatory implications of generative AI

The characteristics described above create several implications for law and governance.

First, as detailed in part 5, companies using generative AI are subject to the same set of legal obligations that pertain to the use of any technology in Australia. At this early stage in the development of generative AI, regulators may be unsure how to enforce these laws. This does not mean those laws are inapplicable to such systems.

Second, the wide range of use cases to which generative AI systems can be put suggests that public moratoria, bans or technology-focused risk categorisations that entirely restrict generative AI use will be ineffective. They will also tend to slow innovation and inhibit beneficial applications. Similarly, organisational-level restrictions need to be carefully communicated and enforced.

Risk-based approaches to regulating generative AI would be more effective when linked to specific use cases, rather than the nature of the technology itself. As HTI has proposed previously in relation to other flexible and powerful AI technologies, including facial recognition technologies, such an approach may well allow for purpose limitations for certain high-risk use cases, while supporting experimentation for genuine research and a wide range of low risk use cases.⁷³

Third, the flexible, scalable, and fluent nature of many generative AI systems suggests that norms and standards around transparency of use will be critical to minimising harms. Such transparency could include notification for individuals interacting with such systems; clear attribution of text, images, sound, or video produced by generative AI; and transparent acknowledgement around issues such as the ownership of and responsibility for such outputs.

Fourth, the opacity and flexibility of generative AI systems pose challenges to accountability. Regulators and corporate policies should carefully attribute legal liability for harms or errors to entities across the AI value chain and incentivise the creation of effective safeguards at the most effective and appropriate points.⁷⁴ Such attribution should note that, particularly in enterprise software, there is significant interplay between the developers (those creating and distributing AI systems), deployers (those offering services powered by AI), and users (individuals, teams or organisations using AI to create content). For example, deployers of LLMs may unilaterally ‘fine-tune’ base models, developers may co-develop customised models with deployers, and users may actively seek to circumvent safety systems.

As an example, developers could be required to provide transparency and security around data flows and model limitations. Deployers could be asked to anticipate malicious or harmful use cases and make efforts to prevent users from circumventing limits. Users could be held responsible for the legal consequences arising from their use of generative AI, while being required to publicly declare where and when content is created. Finally, affected individuals (including recipients or consumers of AI generated content, advice or related outputs, or those misrepresented in outputs) should be made aware of the role that generative AI has played, be given the opportunity for free, informed and prior consent, and be provided opportunities for redress if harms emerge.

Finally, given their multifaceted and complex nature, corporate leaders should seek to create safe spaces – such as secure, organisational platforms – in which thoughtful, yet innovative, experimentation with AI systems can occur. Such spaces should be complemented with clear limits on where the use of generative AI is inappropriate and the full consent of participants. Recent surveys and evidence from HTI interviews suggest that up to a third of professionals have experimented with generative AI in their work.⁷⁵ In the absence of guidelines for such use, organisations could be exposed to risks related to data security, brand damage, or claims for IP infringement.

Part 4.

How are organisations currently governing AI systems?

4.1 The need for AI-focused corporate governance

Corporate governance is ‘the framework of rules, relationships, systems and processes within and by which authority is exercised and controlled in corporations’.⁷⁶ It is ultimately about making effective decisions and accountability. Good governance is therefore a critical organisational asset – it is ‘the framework that ensures the organisation can meet its mission’.⁷⁷

AI should be considered a strategic, enterprise-wide issue. AI systems can be a major enabler or detractor to achieving strategic objectives, and often introduce additional complexity to the operational structure and boundaries of organisations. They must therefore be managed through effective corporate governance principles, structures, and processes.

While still relevant to AI systems, traditional IT governance arrangements are inadequate to address their unique risks and characteristics. AI systems use data – which is often more personally sensitive and externally sourced – at a greater scale than other IT systems. Companies therefore must understand and carefully manage data assets so that poor-quality data doesn’t translate to poor-quality outputs.

Compared to other IT systems, AI systems can be more dynamic, display emergent capabilities, be non-deterministic and act in unintended ways. They are often less transparent than programmed IT systems, which makes testing and validation more difficult. It may not be immediately clear that an AI system is underperforming or performing in inappropriate ways. AI systems are also more likely to be deployed at scale in complex, stakeholder-facing contexts. This means that organisations can face systemic, rather than localised, risks from AI systems.

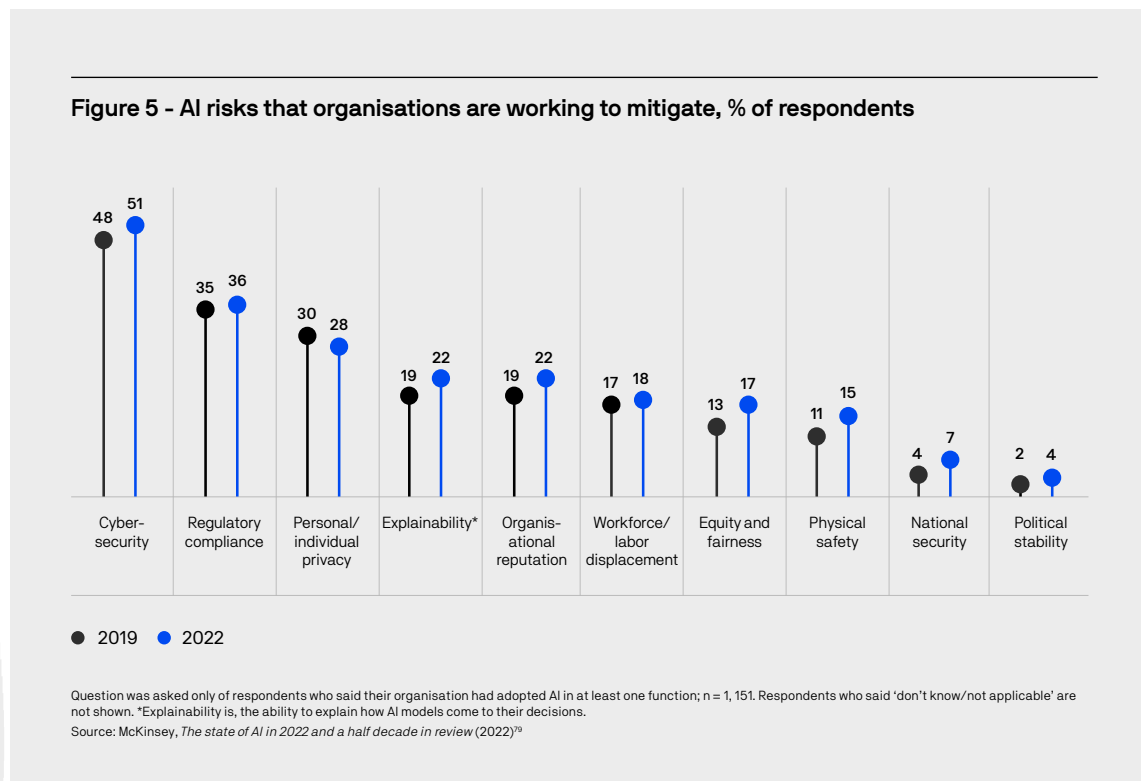


AI should be considered a strategic, enterprise-wide issue.

4.2 Current approaches to AI governance

HTI research reveals that few Australian organisations have implemented systematic and structured forms of governance around AI systems. This finding is supported by other surveys: the 2021 Responsible AI Index found that most sectors in Australia are currently in the initiating phase, with manufacturing and banking sectors identified as being the most advanced.⁷⁸

This is not a uniquely Australian problem. Globally, despite rapid increases in adoption over the last three years, there is little evidence of rising investment in how organisations are practically mitigating AI-related risks (Figure 5). This suggests a growing and concerning AI governance gap.



There is huge in diversity in how Australian organisations manage and govern AI systems.

HTI's qualitative and quantitative research reveals that strategic investment and governance decisions around AI systems are currently made without the benefit of a systematic consideration of the risks of AI systems described above.

Australian organisations are governing AI systems with a variety of approaches that tend to be fragmented, strategically disconnected, blind to the idiosyncratic characteristics of AI and fail to account for potential impacts on stakeholders.

HTI surveyed 268 company directors and senior executives of Australian firms. HTI also held a series of workshops diving into the experiences and perspectives of 50 company directors and conducted individual interviews with 30 corporate leaders in organisations currently using AI.

The quantitative survey revealed that only 10% of corporate leaders indicated that their organisation possesses an AI strategy. A slightly larger percentage – 14% for executives and 13% for company directors – indicated that they have a set of AI or data ethics principles.

Regarding AI systems and processes, 46% of all survey respondents currently using AI claimed that their organisation undertakes a risk assessment for their organisation's use of AI. Just over a third claimed the same is done for their suppliers. 35% of company directors claimed that AI risks were on the risk register, while 12% of senior executives asserted this. 43% of corporate leaders were somewhat or very confident in the skills and capabilities of key personnel to drive AI governance improvements.

Follow-up interviews with respondents and HTI workshop discussions suggest that these figures are highly optimistic. The qualitative nature of the data emerging from these engagements is indicative yet revealing.

Out of 80 corporate leaders asked directly about their governance processes in workshops or interviews, only four participants indicated that their organisation had implemented a structured governance system specifically oriented toward AI models and systems. Approximately a quarter of senior executives and three-quarters of company directors engaged in HTI interviews were unaware or unsure of the details of AI governance in their organisation. Aside from a subset of technology executives, only one interviewee could confidently state that their organisation had a comprehensive view of all the AI systems used in their business.

The interviews reveal that, in practice, there is huge diversity in how Australian organisations manage and govern AI systems. Many of the largest and best-resourced organisations currently conduct AI system oversight via a mix of ad hoc practices and pre-existing structures.

For example, many corporate leaders reported that design, development, and procurement decisions are heavily influenced by a single 'guru', perceived as the organisational expert in AI. At least two large organisations reported that AI-related risks were managed via simple spreadsheets, separate from other IT governance systems. A number of interviewees said that their AI systems were approved by legal and compliance teams with little to no practical knowledge of AI systems.

The shortcomings in current governance approaches are ably illustrated by the elements that corporate leaders believe would help them govern AI systems better.

Only 10% of corporate leaders indicated that their organisation possesses an AI strategy.

As shown in Figure 6, company directors place the need for an up-to-date AI strategy as the highest priority for good governance. Both directors and executives agree that strategic expertise across the organisation, higher levels of risk awareness, documented policies and practices, and access to examples of effective models employed by peers are critical to effective oversight. Executives also saw a role for better regulatory guidance around AI-related accountability as essential.

Corporate leaders should think broadly and strategically when considering how to implement effective AI governance in their organisation. Part 7 of this report suggests several actions corporate leaders can take to address these shortcomings and implement more effective AI governance systems.

Figure 6 – Top five supports required to implement effective AI governance

Company directors	Executives
 Creation or updating of an AI strategy.	 Greater strategic expertise around the use of AI across the organisation.
 Greater strategic expertise around the use of AI across the organisation.	 Greater executive awareness of the potential risks posed by different AI use cases.
 Greater board awareness of potential risks posed by different AI use cases.	 Documentation of policies and practices to identify and mitigate AI risk.
 Documentation of policies and practices to identify and mitigate AI risk.	 Examples of effective AI governance models in peer organisations.
 Examples of effective AI governance in peer organisations.	 Better regulatory guidance around AI-related accountability.

Box 3: The promise and shortcomings of governance via AI principles

HTI interviews revealed that a growing number of Australian organisations are looking to develop and adopt responsible or ethical AI principles to reassure their staff and customers that they are attuned to the potential harms of their AI systems.

Ethics and principle-based approaches can provide a valuable framing for examining the intention, design, and potential consequences of emerging technology as a reflection of organisational values. However, there is emerging evidence that merely developing or adopting a set of ethical principles is not sufficient as an AI governance strategy.

Empirical research indicates that AI principles and codes of ethics have little discernible impact on the behaviour of engineers developing AI systems.⁸² Furthermore, the expression, interpretation and implementation of ethical principles leave them prey to being meaningless ('but what does 'fair' really mean?'), isolated ('I don't think they apply to us') and toothless ('those are just guidelines').⁸³ Moreover, other research indicates that poorly-designed normative frameworks that rely on moral encouragement can be counterproductive, inducing precisely the behaviour the organisation wanted to avoid.

The ineffectiveness of AI principles on their own is supported by experience in other areas of corporate governance. Evidence from the work health and safety field indicates that regulations, inspections, prosecutions, guidance material, campaigns and enforceable undertakings are the most important incentives for businesses to adopt practices that keep workers safe.⁸⁵ Relying on voluntary action, public pressure or outside incentives has little impact on behaviour.

Furthermore, the creation and adoption of AI ethics principles may lead to a form of 'ethics washing' that creates risks for organisations and society.⁸⁶ For adopting organisations, such principles may induce a false sense of security that the problems have been managed. For customers and members of the public, principle-focused statements that are not supported by structured governance systems can raise expectations in ways that result in widespread loss of trust in the ability of organisations to safely use AI systems. This is particularly true when such systems cause harms that directly contradict the public assurances of their ethical and responsible nature.

Naturally, the solution is not to jettison attempts at codifying AI principles. Rather corporate leaders should view such statements as an important base from which to diligently explore and construct the practical governance strategies, policies and institutional structures required to give effect to those principles.

Part 5.

Legal obligations of Australian organisations using AI

Corporate leaders must understand the laws governing AI systems to meet their individual and organisational obligations.

Australia does not yet have AI-specific laws. Instead, the development and use of AI are regulated primarily by technology-neutral laws of general application. Some of these obligations apply to the organisation, while others apply personally to directors, senior executives, and other key personnel.

Unfortunately, the extent and nature of existing legal applications applying to AI systems are not well understood across corporate leaders responsible for developing and deploying AI systems today. Despite increasing enforcement activity by regulators, HTI research reveals that corporate leaders tend to see the lack of AI-specific regulation in Australia as indicative of an ‘AI Wild West’.

This part explores the links between harms that can arise from AI use and the existing Australian laws that may apply to directors and organisations.⁸⁸ This section is not comprehensive, nor should it

be considered legal advice. Organisations may face expanded or additional requirements based on industry sector (for example, organisations operating in the financial sector, organisations delivering essential services such as energy or telecommunications, or public sector organisations), or specific use cases (for example, deployment within a medical or workplace setting).

This part raises questions for corporate leaders to ask in board discussions, during executive meetings and throughout governance processes. These questions may apply at different stages of, or repeatedly throughout, the lifecycle of an AI project, and may need to be considered on an ongoing or periodic basis as AI systems evolve and are updated.

Figure 7 highlights selected obligations for corporate leaders and organisations regarding the design, development and use of AI systems. Table 3 provides an overview of the harms that can arise to individuals in AI use and the existing laws that may apply.⁸⁹ Parts 5.1 and 5.2 go into more detail, beginning with the legal obligations of company directors, then expanding to duties that apply to organisations using AI.

Australia does not yet have AI-specific laws. Instead, the development and use of AI are regulated primarily by technology-neutral laws of general application.



Table 3 – Common harms to individuals and existing laws that may apply

When an AI system...	Existing laws may apply...
Misuses data or personal information	<ul style="list-style-type: none"> • Privacy laws • Data-security obligations • Security of Critical Infrastructure Act • Risk management obligations • Confidentiality obligations • IP laws
Produces an incorrect output	<ul style="list-style-type: none"> • Australia Consumer Law - product liability (if the organisation is a manufacturer) and consumer guarantees • Privacy laws, if the output is personal information
Provides misleading advice or information	<ul style="list-style-type: none"> • Australian Consumer Law - misleading and deceptive conduct, unconscionable conduct, false and misleading representation, consumer guarantees
Provides unfair or unreasonably harsh treatment	<ul style="list-style-type: none"> • Australian Consumer Law - unconscionable conduct • Australian Consumer Law - consumer guarantees
Discriminates based on a protected attribute	<ul style="list-style-type: none"> • Anti-discrimination laws
Excludes an individual from access to a service	<ul style="list-style-type: none"> • Anti-discrimination laws, if the exclusion relates to a protected attribute • Essential service obligations (e.g. electricity hardship and disconnection obligations) • Australian Consumer Law - unconscionable conduct
Restricts freedoms such as expression, association or movement	<ul style="list-style-type: none"> • Human rights acts or charters in Victoria, Queensland and ACT
Causes physical, economic or psychological harm	<ul style="list-style-type: none"> • Negligence, if there is a breach of a duty of care that causes harm • Work, health and safety laws • Australian Consumer Law - product liability (if the organisation is a manufacturer) and consumer guarantees

Figure 7 – Key obligations in the design, development and use of AI

Directors' duties



The legal and regulatory environment

5.1 Duties of company directors related to the use of AI

A company director has duties under the common law and the Corporations Act relevant to their company's use of AI.



Directors' duties regarding AI

Company directors have a common law fiduciary duty to act in the company's best interests and to exercise independent, informed judgement in managing a company. For companies regulated by the Corporations Act, this duty is reinforced by section 181.⁹⁰ Section 180 provides that directors (and other officers of a corporation) must exercise reasonable care and diligence in carrying out their duties and exercising their powers.⁹¹

To satisfy these duties, directors must be able to properly guide and monitor management of the company and make decisions based on an appropriate degree of knowledge of the business of the organisation and its key business risks (including non-financial risks).

These legal duties apply to a director's decision making and oversight of the company's development and use of AI.

Section 180 establishes an objective standard for the degree of care and diligence required by directors. It asks what a reasonable director would have done considering the circumstances of the company and the responsibilities of the relevant director. As AI systems are increasingly relied on by a company in its operations, management of the risks of AI become increasingly important.

For example, there has been greater focus on this duty of care and diligence in the context of governance failures in meeting cyber security and privacy obligations.⁹² Given the increased threat of cyber attacks, effective governance in managing any applicable cyber security risks now involves active engagement by directors.⁹³ A company's use of AI raises both cyber security and privacy risks, as well as other harms and risks.

Company directors must exercise reasonable care and diligence to ensure that there are appropriate oversight and governance systems for AI systems relied on by their company.

To discharge their duties when companies are using AI, directors must have a sufficient understanding of both the business risks and the law that applies to their company and its use of AI. Under the 'stepping stone' doctrine, a director may be liable under their statutory directors' duties for failing to take reasonable steps to prevent foreseeable harm resulting from serious breaches of the law by the company.⁹⁴ The foreseeable risk of harm to the corporation is not just financial, but also to its reputation and its interests in complying with the law.⁹⁵

Under Australian common law and s181 of the Corporations Act, directors must consider what are the best interests of the company. They can, and arguably should, consider how stakeholders, such as employees, customers, suppliers and the local community, will be affected by the company's use of the AI, given the company's interests in avoiding reputational harm.⁹⁶

Table 4 – Directors' duties and the use of AI

Corporations Act 2001 (Cth)

Due care and diligence (s180) Directors and officers must exercise their powers and discharge their duties with the degree of care and diligence that a reasonable person would exercise if they were a director or officer of the same corporation, bearing in mind the corporation's individual circumstances, and the office holder's position and responsibilities.

Directors are responsible for ensuring that effective risk management and compliance systems are in place to assess, measure and manage any risks and impacts associated with a company's use of AI.

Act in good faith (s181) Directors and officers must exercise their powers and duties in good faith, in the best interests of the corporation, and for a proper purpose.

For AI systems, this requires directors to be informed about the subject matter and rationally believe their decisions are in the best interests of the company, having properly considered the potential impact of those decisions.

Questions
corporate
leaders should
be asking...



- ❓ Are our governance systems and structures fit-for-purpose to identify, assess and manage the risks posed by AI?
- ❓ Do we have the right resources, capabilities, skills and training on and around the board and executive team to respond? How should we draw on external expertise?
- ❓ Is there a clear accountability framework and appropriate reporting to the board on the strategic opportunities and risks posed by AI to the organisation?

5.2 Legal obligations for Australian organisations using AI

Organisations using AI systems in Australia are subject to various legal obligations. Both directors and senior executives should be aware of how these apply.

As illustrated in Figure 9 above, six areas of law have particular relevance for AI systems: privacy and data, consumer protection, cyber security, anti-discrimination, negligence and work health and safety.⁹⁷

Privacy and data

Personal information (PI) is collected and used to train a wide range of the most popular AI systems used by Australian organisations today. Even where this is not the case, PI is often an input to, or an output of, a deployed AI system. The development and use of AI systems therefore often engage privacy laws, which stipulate requirements for the collection, use, disclosure and transfer of PI in both the public and private sector. Organisations therefore need to be mindful of how the data used in AI systems is collected, managed, used, and

disclosed (whether done by their organisation or a third party). This is especially the case when the organisation collects and uses sensitive personal information, including biometric information.

In Australia, the *Privacy Act 1988* (Cth) (the Privacy Act) and the associated Australian Privacy Principles (APPs) apply to Australian Government agencies, and private sector organisations subject to some exemptions.⁹⁸ State-based privacy and human rights laws are also further sources of privacy-related obligations. While a tort of privacy has not been accepted⁹⁹ or introduced in Australian law, if such a regime was introduced, this would introduced¹⁰⁰ greater civil liability for organisations that undermine privacy.

Entities regulated by the Privacy Act and using or creating PI with AI systems must meet obligations under the APPs (outlined in Table 5).



Table 5 – Privacy obligations applying to AI use

Privacy Act 1988 (Cth)

Open and transparent management of PI (APP 1)	<p>Where AI systems use PI, organisations must ensure that the PI is being managed in an open and transparent manner, including by:</p> <ul style="list-style-type: none"> • having an up to-date privacy policy, which should cover any AI-related use or generation of PI; and • taking reasonable steps to implement organisational practices, procedures, and systems to ensure compliance with the rest of the APPs. <p>This requires organisations to have ongoing governance over their PI use in AI systems. This includes, where AI systems ‘learn’ or develop over time, regular monitoring, and assessment.</p>
Collection of information and data minimisation (APP 3, 5)	<p>Organisations using PI in AI systems must not collect PI, unless reasonably necessary for its functions and, for sensitive information (which includes certain biometric information), with consent (unless an exception applies).</p> <p>Organisations need to ensure PI has been collected lawfully and fairly, and that individuals have been notified of specific matters about the collection, including the kinds of information collected and the purpose of collection.</p> <p>Practically, this means that data minimisation should be considered as part of all AI deployments.</p>
Use or disclosure of personal information (APP 6)	<p>Where organisations have collected PI for a particular purpose, they can only use or disclose that data for a second, unrelated purpose in limited circumstances, such as with consent. This means that organisations must be mindful to ensure that AI systems are not reusing existing PI datasets collected for unrelated purposes.</p>
Cross-border disclosure of personal information (APP 8)	<p>Organisations engaging in cross-border disclosure of PI must take reasonable steps to ensure that overseas recipients (including overseas cloud-based service providers or AI developers) do not breach the APPs in relation to that PI. Liability for the practices of the overseas entity in breach of the Privacy Act in relation to the handling of any PI collected by the organisation often rests with the responsible APP entity.</p>
Quality assurance (APP 10)	<p>Organisations need to take reasonable steps to ensure that the data that the organisation collects, uses, and discloses is accurate, up-to-date, complete and relevant. Organisations will also need to ensure that the outputs of an AI system which creates PI (such as inferences about an individual) meets these quality measures.¹⁰¹ This requires ongoing quality assurance to be applied to the AI systems, both in respect of input data and outputs.</p>

Questions corporate leaders should be asking...



- ❓ Are there processes to ensure the use of PI is minimised and that an AI system's use of PI (as opposed to de-identified data) is reasonably necessary?
- ❓ What systems do we have to ensure that any PI used in our AI systems has been collected, used and disclosed lawfully and fairly?
- ❓ What is the source of the PI that our AI systems are using? Are we confident that this data has been collected fairly and lawfully?
- ❓ How are we ensuring that PI collected for a particular purpose is not being reused for other purposes within our AI systems? Or, if we are using PI for secondary purposes, are we obtaining free and informed consent?
- ❓ What systems do we have in place to ensure that the PI we hold for use by our AI systems is accurate, up-to-date, complete and relevant? If the AI generates inferences about individuals, how are we sure those inferences are correct?

In addition to privacy considerations, an organisation that develops or deploys an AI system must ensure that it has the necessary rights, including IP rights, contractual rights and rights with respect to the use and disclosure of confidential information, to use that data with the AI system.

All data that comes into or is collected by an organisation will likely have confidentiality, IP and/or contractual restrictions that apply to it, whether these are common law or equitable duties of confidentiality, under IP legislation such as the *Copyright Act 1968* (Cth), or via customer terms or other third party or supplier contracts.

Consequently, good data governance and data and privacy provenance is critical in enabling organisations to use data with AI systems without breaching these obligations.

It is important to note that a number of important changes proposed by consumer, privacy and human rights organisations have been reflected in the Attorney-General's Department *Privacy Act Review Report* (2023), such as: modernisation of the definition of personal information; expanded coverage of the Privacy Act to businesses with a turnover of under \$3 million and potential removal of the 'small business' exception; introduction of a direct right of action for breaches; the right to object to collection, use and disclosure and the right to erasure; increased transparency requirements around the use of PI in 'substantially' automated decisions which have a legal or similarly significant effect on an individual's rights; and additional protections for the collection and use of biometric data, such as conducting privacy impact assessments where there are high risks to privacy, or risk assessment requirements for facial recognition technology.¹⁰²



Consumer protection

Consumer protection laws are designed to promote fair trading and competition by establishing rigorous consumer protections that govern the interaction between manufacturers, suppliers, and consumers (including some business-to-business arrangements). These protections enshrine the rights of consumers and companies within marketplaces, aiming to address power imbalances that can arise between consumers and businesses, and between some businesses. These laws also address other issues such as disclosure, fairness, deceptive conduct, defective goods, and responsible lending practices.

Consumer protection obligations mostly derive from the Australian Consumer Law in Schedule 2 of the *Competition and Consumer Act 2010* (Cth) (Australian Consumer Law). Organisations

supplying AI-enabled products or services to consumers¹⁰³ (including the provision of information or advice) are subject to Australian Consumer Law. Other pieces of legislation impose further sector-specific obligations on financial service licence holders, credit licence holders, energy retailers and telecommunications providers with a consumer protection focus.¹⁰⁴

Consumer protection laws could be triggered by the use of AI, for example, by deploying customer-facing AI such as automated decision-making systems and chatbots. Consumer protection laws may also apply where the AI system is the product or service - or part of the product or service - that is being marketed to the consumer.



Table 6 – Australian Consumer Law obligations applicable to AI use

Australian Consumer Law

<p>Misleading & deceptive conduct (s18)</p>	<p>Organisations using AI systems in trade or commerce cannot engage in conduct that is misleading or deceptive, or likely to mislead or deceive. For example, organisations using AI to make decisions cannot suggest that the decision is being made by a staff member, when in fact it was made by an algorithm. This prohibition applies even where there is no intention to mislead or deceive.</p> <p>Directors and individual personnel may be held personally liable for breaches.</p>
<p>Unconscionable conduct (s20-22A)</p>	<p>Organisations using AI systems in trade or commerce must not make decisions or produce outcomes that are so harsh that they go against good conscience. This is important in an AI context given that unlike a human decision maker, AI cannot assess the fairness of an outcome.</p> <p>Directors and individual personnel may be held personally liable for breaches.</p>
<p>False or misleading representations (s29-37)</p>	<p>A person involved in trade or commerce must not make false or misleading representations about goods or services or engage in misleading conduct in respect of these goods and services. As with misleading and deceptive conduct, organisations must not misrepresent when an AI system is used, nor should they misrepresent how the AI system reaches its outputs, or the accuracy and reliability of those outputs. This obligation applies when dealing with consumers and with other businesses.</p> <p>Directors and individual personnel may be held personally liable for breaches.</p>
<p>Product liability (Part 3-5)</p>	<p>Manufacturers of AI systems and AI-enabled ‘goods’ with a ‘safety defect’ (i.e. where the goods are not as safe as people are entitled to expect) that causes particular harms (e.g. personal injury or property damage) are strictly liable to compensate that loss or damage. Manufacturers may also be required to indemnify the ‘supplier’ for any liability under a consumer guarantee where the loss or damage is caused by the safety defect.</p>
<p>Consumer guarantees (Part 3-2 Div 1 Subdivision A and Subdivision B)</p>	<p>Consumer guarantees may apply to organisations supplying AI-enabled systems to consumers to a value of \$100,000 or less, or that are of a kind ordinarily acquired for household use and consumption. For goods, these include the guarantee as to acceptable quality, fitness for purpose, and the guarantee that the goods match the description. An AI-enabled system that produces inaccurate or unconscionable results may be unfit for purpose and not of acceptable quality.</p> <p>If the AI-enabled system is supplied as a service, the guarantee that the service be rendered with due care and skill and be reasonably fit for the consumer’s known purpose will apply.</p> <p>For AI systems that are goods, breaches of consumer guarantees require the repair or replacement of the good, or the payment of the cost of replacing or repairing the goods. For AI systems that are services, breaches of the consumer guarantees require the service to be supplied again, or payment of the cost of having the service supplied again.</p>



Questions corporate leaders should be asking...

- ❓ Where AI systems are providing advice or recommendations to - or making decisions about - consumers, how are we ensuring representations are not misleading or deceptive in nature? How are we ensuring the AI system is not behaving in a misleading or deceptive way?
- ❓ Are we clear with our customers about when we are using AI, and how we are making decisions that may affect them?
- ❓ Do we understand the potential outcomes for consumers impacted by the AI systems we use? Is there the potential for the outcome to be so harsh as to go against good conscience?
- ❓ What processes do we have in place to detect and prevent safety defects in the AI-enabled products and services we supply?
- ❓ Where we are supplying AI-enabled systems to consumers, what testing are we doing to ensure that they are fit-for-purpose, of acceptable quality and match the description (including the AI outputs)? Are we supplying AI services with due care and skill?

In addition to obligations under the Australian Consumer Law, Australian Financial Service Licence (AFSL) holders and Australian Credit Licence (ACL) holders have additional consumer protection obligations under the *Corporations Act 2001 (Cth)* and *National Consumer Credit Protection Act 2009 (Cth)* which may apply to AI use, including product design and distribution obligations, disclosure requirements, and the

supervision of provision of financial product advice and obligations to provide services 'efficiently, honestly and fairly'. In addition, there are various competency requirements which will also apply to organisational competencies with respect to AI systems in use in provision of certain services.



Cyber security

Cyber security is a key consideration for organisations that are developing and deploying AI, given the significant volumes of data often involved and the connectivity of AI systems. Existing data security and data breach obligations include:

- Security, destruction and de-identification of PI, and notification of data breaches under the *Privacy Act 1988*
- Sector- or industry-specific regulation, particularly for:
 - APRA-regulated entities, including various risk management and data security obligations such as CPS 220 (Risk Management) and CPS 234 (Information Security). Significantly, these prudential standards make clear that the board is ultimately responsible for compliance with these requirements.

- Australian financial services licensees have obligations to act efficiently, honestly and fairly, to have appropriate risk management in place and to have adequate financial, technological and human resources¹⁰⁵ which has been applied by ASIC in relation to ensuring adequate cybersecurity measures are in place¹⁰⁶.
- ASX-listed entities have various disclosure obligations, including in response to events that a reasonable person would expect to have a material effect on the price or value of its securities, which can include large cyber incidents.¹⁰⁷
- Reporting, risk management practices, governance assistance and other cyber security obligations for entities regulated by the *Security of Critical Infrastructure Act 2018* (Cth). Telecommunications entities (carriers and carrier service providers) have similar notification and reporting obligations under the *Telecommunications (Carriage Service Provider – Security Information) Determination 2022* (Cth).



Table 7 – Cyber security and data security obligations applying to AI use

Privacy Act 1988 (Cth)

Data security, destruction and de-identification (APP 11)	Organisations with AI systems that are using, collecting or disclosing personal information, must take reasonable steps to protect PI, which includes destroying or de-identifying information that is no longer needed.
Data breach notification (Part IIIC)	If an organisation holding and using PI as part of its AI system is impacted by an eligible data breach, ¹⁰⁸ affected individuals and the OAIC must be notified. ¹⁰⁹

APRA Prudential Standards CPS 234 (Information Security)

Data security obligations (S 13-36)	<p>APRA-regulated entities have a range of detailed information security obligations which apply to data used by AI systems. Boards are responsible for:</p> <ul style="list-style-type: none"> • the ultimate security of the organisation’s data • defining roles & responsibilities for decision making, approval, oversight and other security functions. <p>CPS 234 also imposes requirements in respect of: information security capabilities; implementation of a data security policy framework; information asset identification & classification; implementation of controls; incident management; testing control effectiveness; internal audit; and, APRA notification of material cyber security incidents within 72 hours.</p>
--	--

Security of Critical Infrastructure Act 2018 (Cth)

Responsible entities must have a critical infrastructure risk management program. They also have obligations for reporting of operational information, notification of cyber security incidents, and enhanced cyber security obligations for systems of national significance.



Questions
corporate
leaders should
be asking...

- ❓ What policies and procedures do we have in place to ensure the management of our AI systems - and the data they draw upon – appropriately protect data and meet cyber security & data breach notification obligations?
- ❓ Do we have risk management frameworks in place to identify, manage and mitigate AI cyber risks, and to detect and manage cyber incidents?
- ❓ Have we defined clear roles and responsibilities for cyber security within our organisation, and do we have the right resources (people, capabilities and technology), advisers and training in place, particularly with respect to large data models used by AI?
- ❓ Have we considered the additional cyber risk associated with retention and use of large data models?
- ❓ What testing and internal reviews are we conducting to give us confidence that organisational controls and policies are operating as intended?



Anti-discrimination

The purpose of anti-discrimination legislation is to uphold equality and fairness by prohibiting discrimination based on certain grounds such as race, sex, ability, and age. Federal legislation – most notably, the *Racial Discrimination Act 1975* (Cth), *Sex Discrimination Act 1984* (Cth), *Disability Discrimination Act 1992* (Cth), and *Age Discrimination Act 2004* (Cth) – operate concurrently with state and territory anti-discrimination laws.

The risk of bias and discrimination associated with AI systems is well documented,¹¹⁰ and it is therefore imperative that organisations implementing AI systems actively monitor these systems and their outputs for bias or unfairness which may result in discrimination based on protected attributes, such as a person's age, disability, disability carer status, race, colour, descent, national or ethnic origin, immigrant status, sex, sexual orientation, gender identity, intersex status, marital or relationship status, pregnancy status, breastfeeding or family responsibilities.

Table 8 – Anti-discrimination obligations applying to AI use

Anti-discrimination laws¹¹¹

Anti-discrimination laws prohibit organisations from using an AI system that directly or indirectly discriminates against people with protected attributes.

Organisations must monitor AI systems to ensure that they do not produce biased outputs. This will involve a consideration of the data used to train and drive the AI system and any bias within that training data. This is also an ongoing requirement as the way AI systems develop and 'learn' may start to reinforce and amplify bias over time.

Organisations must also consider the accessibility of their systems and make reasonable adjustments to ensure that people with disabilities are not disadvantaged when engaging with the AI system.



Questions corporate leaders should be asking...

- ❓ What processes do we have in place to ensure that the data being used to train or drive our AI systems are sufficiently broad, large, diverse and not affected by bias?
- ❓ Do we have diversity in the development teams?
- ❓ Have we assigned roles within the organisation to identify and mitigate bias in our AI systems?
- ❓ What audit and monitoring systems do we have in place to ensure there is not bias in the ongoing use of the AI system?
- ❓ If our AI systems engage with consumers or other individuals, what steps are we taking to ensure that they are accessible to everyone and do not disadvantage members of protected groups?



Duty of care and negligence

Organisations using AI systems may have a common law duty of care towards people that use or are impacted by the system.

The law of negligence, codified in some state-based civil liability schemes, provides that if an organisation owes a duty of care to a class of persons, it must exercise the standard of care of a reasonable person in the circumstances to avoid foreseeable injury or loss to the relevant persons, and may be liable for loss or injury suffered by those persons where the organisation fails to exercise that standard of care. The duty can be imposed on manufacturers, retailers, and distributors of

AI systems, and in certain specific relationship scenarios where AI systems are used, e.g. doctor to patients, employer to an employee.

The tort of negligence has a broad application (and is constantly evolving) and could extend to a number of use cases of the outputs of AI systems, including advice, economical, physical and psychological injury, industrial outputs and administrative decision-making.

In addition, where a duty of care is owed to a person and an AI-enabled product creates loss or damage, both product liability (see Table 6) and negligence claims may apply.

Table 9 – Duty of care and negligence and their application to AI products and use

Duty of care and negligence¹²

Organisations that have a duty of care to a class of persons:

- must exercise the standard of care of a reasonable person in the circumstances to avoid foreseeable injury or loss to the relevant persons;
- may be liable for loss or injury suffered by those persons where the organisation fails to exercise that standard of care; and that failure caused a person loss or damage.

In particular, manufacturers, retailers, distributors and donors have a positive duty to exercise reasonable care. The extent of this duty would depend on the level of risk associated with the activity in question. In the case of machine learning systems, that duty may involve sufficiently testing, monitoring and/or supervising of the system.

- ❓ What processes do we have in place to identify and mitigate potential harms that may arise to people who interact with an AI product or system?
- ❓ Have our AI systems been sufficiently designed, tested, developed and monitored to avoid foreseeable harm?
- ❓ How are our AI systems captured within our organisations risk assessment framework?



Questions corporate leaders should be asking...



Work health and safety

Deployment of AI systems within a workplace context can introduce risks of physical and psychological harm to employees. Work health and safety ('WHS') laws require that:

- Organisations ensure, as far as reasonably practical, the health and safety of workers and other persons. This will include factoring AI into health and safety training;
- Corporate leaders must also exercise due diligence to ensure organisations meet their WHS obligations.

Table 10 – Work health and safety obligations that apply to AI use

Work health and safety laws¹¹³

Corporate leaders must exercise due diligence to ensure organisations meet their WHS obligations. Deployment of AI systems within a workplace context can introduce risks of physical and psychological harm to employees.

Workplace laws require that organisations ensure, as far as reasonably practical, the health and safety of workers and other persons, including factoring AI into health and safety training.

Questions
corporate
leaders should
be asking...



- ❓ How do our WHS framework and safety procedures incorporate and address any workplace safety risks arising from the use of AI systems?
- ❓ Are AI systems that may impact workers being included in health and safety training?



Part 6.

International trends in AI law and policy

Australia lags internationally in the design and implementation of AI-specific regulation. Many jurisdictions are reforming their laws to ensure that AI systems are deployed in line with local values and public expectations. New obligations are being introduced by amending existing laws and proposing AI-specific laws.

As a critical component of strategic foresight, Australian corporate leaders should be aware of these international regulatory trends and their potential impact on their organisations. Understanding the evolving regulatory landscape outside of Australia is essential for companies doing business overseas. For domestically-focused organisations, international regulatory trends provide important insight into how Australian law may evolve. Understanding these issues will also help Australian organisations remain competitive with global peers.

6.1 The global trend towards AI regulation

Governments and the private sector agree that reform is needed to manage the risks of AI systems while continuing to encourage innovation.

In April 2023, G7 digital ministers agreed that their countries should adopt risk-based regulation on AI, strive for interoperability across AI governance frameworks and support the creation of technical standards for implementation.¹¹⁴

Industry leaders are also calling for regulation. A 2023 US Chamber of Commerce report stated that ‘a failure to regulate AI will harm the economy, potentially diminish individual rights, and constrain the development and introduction of beneficial technologies.’¹¹⁵ CEOs and senior executives from technology companies including Microsoft,¹¹⁶ Google,¹¹⁷ IBM,¹¹⁸ OpenAI¹¹⁹ and Salesforce¹²⁰ have all called for AI-focused laws.

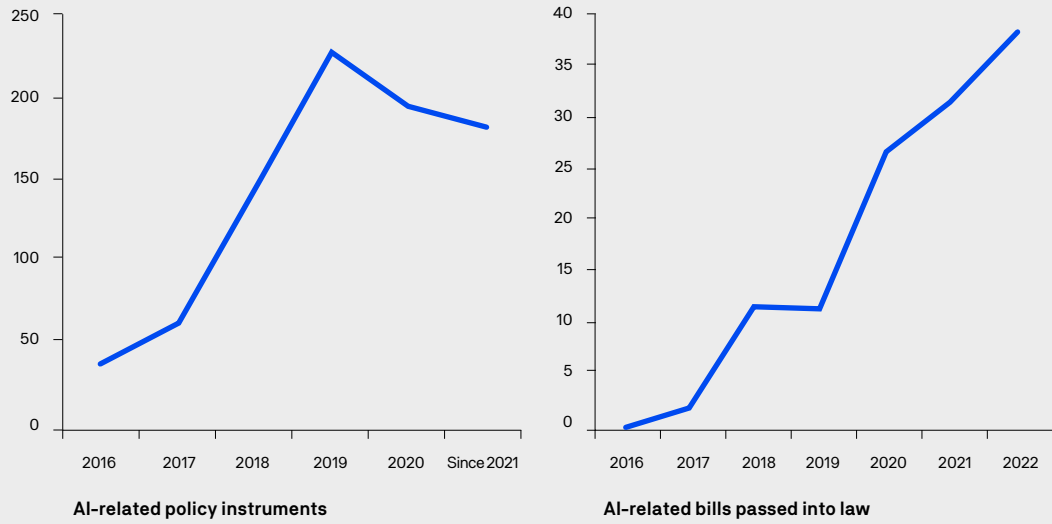
Policy makers across the world are responding. Since 2016, the number of AI-related policy instruments¹²¹ and laws¹²² introduced globally has increased significantly each year (Figure 8).¹²³ Europe has introduced the most AI policy instruments (Figure 9).¹²⁴

Perhaps the most notable international example of AI-specific regulation currently under development is the European Union’s AI Act.¹²⁵ The Act proposes a risk-based approach to ensuring that AI systems are overseen by people, are safe, transparent, traceable, non-discriminatory, and sustainable.¹²⁶ It classifies AI systems into four levels of risk, from minimal to unacceptable.

The progress of the EU Act is being closely watched by policy makers in other jurisdictions worldwide. If the AI Act comes into force, it will impact organisations operating outside the EU through at least three channels: market access, as non-EU organisations will have to comply with EU rules to access the common market; standardisation, as the EU Act will rely on yet-to-be-determined harmonised standards for much of its compliance, which are likely to incorporate or mirror international standards currently under development; and regulatory cooperation, as the Act will encourage the EU’s trading partners to coordinate and align on interoperable AI rules, including via multilateral and bilateral agreements.¹²⁷

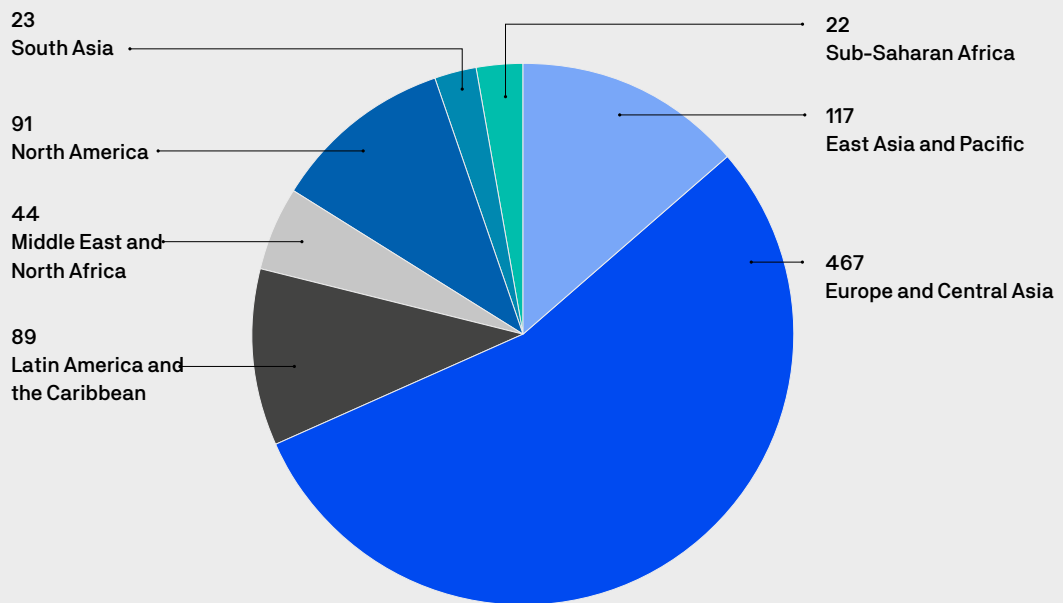
Since 2016, the number of AI-related policy instruments and laws introduced globally has increased significantly each year.

Figure 8 – AI-related policy instruments and laws



Source: HTI analysis based on data from OECD.AI (left chart)¹²⁸; Stanford HAI, AI Index Report 2023 (right chart).¹²⁹

Figure 9 – Policy instruments by region¹³⁰



Source: HTI analysis based on data from OECD.AI¹³¹

Australian companies that operate overseas generally must comply with the laws of the jurisdictions in which they operate. In addition, international regulatory trends impact Australian organisations in three ways:

First, the globalised nature of digital markets, technology and supply chains means that foreign laws can have an extraterritorial application. For example, *Europe's General Data Protection Regulation* (GDPR) has already had an extraordinary influence on website design worldwide, because many companies find it inefficient to create two websites – one for their European operations and one for other markets. A similar effect can be anticipated from the EU's AI Act.¹³²

Second, as discussed in part 6.3, evolving legal norms for the development and use of AI are being influenced by and becoming the foundation for the development of international technical standards on AI.

Third, while Australia has not been quick to enact AI-related reform, international regulatory efforts will inevitably influence domestic regulatory efforts. State and federal governments are actively discussing reform in the development and use of AI, especially as it bears on privacy, automated decision making, and generative AI.

In March 2023, the Hon. Ed Husic, Australia's Minister for Industry and Science, directly linked government discussions on AI regulation to the effectiveness of corporate governance of AI. He said, 'If businesses don't get their frameworks right, the community expectation – absolutely understandably – is that governments will step in. Better to think ahead and get it right that way.'¹³³

6.2 Key features of international AI initiatives and interventions

In most jurisdictions, AI policy and law are underpinned by a set of governing principles for responsible or ethical AI use. For example, in 2019, Australia's Department of Industry, Science and Resources published a set of eight voluntary AI ethics principles designed to guide businesses to implement AI systems responsibly.¹³⁴ A comparison of the ethical principles in Australia and other jurisdictions appears in the Appendix. Common principles include: safety; transparency; explainability; fairness and equality; accountability and oversight; dispute resolution; right to object; privacy; and security. While voluntary, the Australian AI ethics framework is among the most comprehensive and includes all such principles.

Globally, policy makers are increasingly intervening with regulation to ensure that AI is being designed, developed, and used in ways that align with local values and public expectations.

The goals of jurisdictions introducing regulatory intervention strongly align with common AI principles of responsible and ethical AI use identified in the Appendix. Other shared objectives of policy makers include:

- Ensuring legal certainty and enabling effective enforcement
- Incentivising organisational governance
- Supporting innovation and the responsible adoption of AI systems
- Minimising risks to employees, customers and citizens
- Protecting human and consumer rights

Overall, international trends in AI reveal three important patterns:

Globally, policy makers are increasingly intervening with regulation to ensure that AI is being designed, developed and used in ways that align with local values and public expectations.

First, regulatory initiatives and interventions around the world and in Australia strongly suggest that AI-related obligations will emerge from a variety of sources. These include the more rigorous enforcement of existing laws; reform processes that more clearly bring AI-related systems and risks into existing regulatory systems; and new interventions targeting specific AI technologies, use cases and harms.

Second, defining what constitutes an AI system is proving an important point of regulatory debate. A narrow definition of AI systems may allow developers and deployers to escape oversight through technical means. On the other hand, a broad definition may capture software systems that are inherently less risky.

Third, many AI-related regulatory proposals combine elements of a 'risk-based approach' upon the foundation of a 'rights-based approach'. This means that forthcoming regulation in Australia may be based on fundamental rights and freedoms, while differentially regulating AI systems depending on the limits they may pose to these risks.

These trends highlight the need for corporate leaders to stay abreast of regulatory changes across various jurisdictions, while ensuring that organisational governance systems and practices are designed to balance commercial goals with the interests of employees, customers, and other stakeholders.

These trends highlight the need for corporate leaders to stay abreast of regulatory changes across various jurisdictions.



6.3 The role of standards in AI governance

Both Australian and international standards are an increasingly important tool for AI governance and regulation. As Standards Australia puts it, Standards are voluntary documents that set out specifications, procedures and guidelines that help to ensure products, services, and systems are safe, consistent, and reliable.¹³⁵ In addition to codifying leading practices, they support the interoperability of practices and policies across jurisdictions.¹³⁶

Policy makers often leave prescriptive requirements in technical areas to national or international standards. This is the case for the EU's AI Act, which relies on harmonised standards to provide technical solutions to providers to ensure compliance with the regulation.¹³⁷

Such standards are traditionally developed through consensus processes involving technical experts from academia, business, and government. This practical co-regulatory approach can promote technical accuracy. For example, the EU's draft AI Act relies on forthcoming harmonised standards to specify the governance and risk management measures that organisations must comply with to conform with the Act.¹³⁸

Standards are not perfect. In particular, standards may not sufficiently recognise or protect human rights and other public interests. This may be due to a need for related expertise or a structural dearth of input from civil society in standard development bodies.¹³⁹ As with other policy making processes, standards are often shaped most directly by individuals and organisations who possess the necessary resources to participate over extended periods. This means that standards, while expert-based and less politically-biased than many other regulatory approaches, tend to under-represent input from civil society groups, marginalised communities, non-Western perspectives and smaller organisations.

The AI standards most relevant for corporate leaders today are an evolving set of management standards focused on governance, risk management and impact assessment. Some standards are yet to be finalised and published, while others have not yet been broadly adopted. Standards of note include:

- NIST's AI Risk Management Framework (AI RMF 1.0) (published February 2023), a voluntary framework produced by the US National Institute of Standards of Technology that aims to 'improve the ability to incorporate trustworthiness considerations into the design, development, use, and evaluation of AI products, services, and systems.'
- The International Standards Organisation (ISO) standard 22989 (published in July 2022), which provides an overview of AI terminology and establishes a common baseline for AI terms and concepts.
- ISO 23894 (published February 2023), which offers strategic guidance to organisations across all sectors for managing risks connected to the development and use of AI.
- ISO 38507 (published April 2022), which provides guidance for governing bodies such as boards to enable and govern the use of AI.
- IEEE P2863 Recommended Practice for Organizational Governance of Artificial Intelligence (forthcoming), specifies governance criteria and steps for effective implementation for the development or use of AI within organisations.¹⁴¹ IEEE has other relevant AI systems standards in its 7000 series.
- ISO 42001 (expected 2023) and ISO 42005 (forthcoming), which detail standards for AI risk management systems and AI impact assessments respectively.

Part 7.

Actions for corporate leaders

To avoid harms, manage risks and capitalise on opportunities, corporate leaders should take urgent action to improve the governance of AI used by their companies.

AI systems are penetrating the core of business operations. Yet, HTI research indicates that most corporate leaders do not have visibility over how their organisations use or rely on AI. Few organisations have governance systems attuned to the specific challenges posed by AI systems, and corporate leaders report a shortfall in AI-related skills. This makes it difficult, if not impossible, to ensure compliance with applicable law.

The Australian public is distrustful of AI use by businesses and governments alike. This distrust is largely warranted. There are many positive use cases for AI, and the thoughtful and responsible use of low-risk systems is essential to driving productivity increases and delivering better services. However, when AI systems perform poorly, are used in misleading or malicious ways, or are deployed inappropriately, both organisations and affected individuals face significant risks.

There is no ‘wild west’ of AI regulation. On the contrary, regulators across Australia are starting to rigorously enforce a wide range of current, technologically neutral laws that apply to AI systems. Meanwhile, policy makers are considering new forms of AI regulation. Both of these facts should spur corporate leaders to develop AI governance approaches that recognise current obligations in ways that increase the trustworthiness of their systems.

HTI research indicates four areas where corporate leaders can simultaneously discharge their obligations and grasp the opportunities of AI. These are:

1. developing strategic capacity related to AI
2. creating a fit-for-purpose AI strategy
3. implementing AI-specific governance systems
4. setting a human-centred culture around AI

Table 12 provides questions designed to assess the current state of their organisation in relation to AI governance and help company directors and senior executives execute these actions.

AI systems are penetrating the core of business operations. Yet, HTI research indicates that most corporate leaders do not have visibility over how their organisations use or rely on AI.



Action 1: Build capacity and develop strategic expertise in AI

Corporate leaders should engage in both personal and organisational capacity building related to AI. This should extend beyond the development of technical skills and capabilities.

The pervasive and essential nature of AI systems demands that corporate leaders develop a broad awareness of AI systems, know how and why AI is being used in their organisations, and understand the benefits, harms, risks, and legal obligations related to AI.

Corporate leaders should also invest in strategic expertise in relation to AI across the organisation. While necessary, acquiring technical data science skills and capabilities is not sufficient. Corporate leaders, operational teams and front-line staff need a ‘minimum viable understanding’ of why AI systems are being used, how they operate, and what is required to deploy and manage them lawfully and safely. To do this well, organisations will need to adopt a comprehensive and cross-functional approach to AI skills development and capacity building.

Action 2: Create a fit-for-purpose AI strategy

Few Australian organisations currently possess a strategy detailing why they want to adopt AI, where it could add the most value, their risk tolerance, and how they can responsibly experiment to prove its usefulness.

As AI systems penetrate the heart of business models, corporate leaders should ensure their organisations have an AI strategy that clearly expresses the overarching business case for investing in AI systems and how organisational values are expressed through the use of AI. The strategy should set expectations, prioritise opportunities, recognise potential harms and risks, acknowledge legal obligations, and communicate the risk

appetite for AI deployment. Importantly, strategy should be closely aligned to broader organisational objectives and complement existing policy frameworks and risk and assurance practices.

Given the fast-moving nature of AI research and the rapidly expanding market for AI solutions, an AI strategy should recognise that new opportunities and risks may emerge, and levels of organisational risk appetite may change over time. The AI strategy should also support the development of a human-centred culture around the development and use of AI (in line with Action 4).

Action 3: Implement an integrated, comprehensive AI governance system

Organisations need to adopt a cross-functional approach to the management, governance, and deployment of AI systems.

Corporate leaders should ensure that their organisations develop and implement an integrated, structured, and comprehensive governance system for all AI systems used by or for the benefit of their organisation. Such a governance system should be integrated, or at least interoperable with, the organisation's existing risk management processes. Importantly, policies and practices should cover and account for authorised, internally managed AI systems, AI systems deployed or managed on behalf of the organisation by third parties, AI systems that are used in the organisation's supply chain, and 'shadow' or unauthorised AI systems used by employees or contractors.

The vast majority of harms to individuals resulting from AI are foreseeable – indeed, they are being experienced around the world already. Similarly, many of the commercial, reputational, and regulatory risks that AI systems pose to organisations are already evident.

At a minimum, a governance system should:

- a. Establish the processes and policies for procuring, developing, or using AI systems, how these intersect with other policies, and any constraints or boundaries (such as a policy not to use facial recognition systems).
- b. Identify, describe, and document each AI system used across the organisation, including the intended purpose and outcomes, desired benefits, type of AI used, context and scope of use, stage of implementation, sources of data, identified harms and risks, any controls or systems in place to mitigate risks, and other relevant information. These processes should help to improve the explainability of AI systems.
- c. Provide a structured means of identifying and documenting the possible impacts of each AI system on stakeholders, including potential benefits, harms to individuals, and risks to the organisation.
- d. Determine the legal requirements or obligations applicable to each AI system. In addition to identifying obligations to external regulators and stakeholders, the governance system should detail the policies applying to the use of AI systems by employees, contractors, and suppliers.
- e. Establish appropriate oversight of and responsibility for each AI system. Clear lines of delegation and accountability should exist that apply to AI system failures, malicious use, or overuse. The governance system should establish adequate human oversight and ensure that those responsible for using AI systems have the necessary understanding, information, training, control, and authority to make decisions.

Action 4: Set a human-centred AI culture

Corporate leaders should support the development of a human-centred culture around the development and use of AI.

Like all technological systems, AI systems encode fragments of organisational culture at multiple points in the AI life cycle.¹⁴² When AI systems engage with stakeholders or make consequential decisions, they mirror these values, making them powerful expressions of culture.

Corporate leaders should therefore be explicit about their organisational values and realistic as to how these are expressed in AI systems. They should ensure that AI systems deliver appropriate value to stakeholders and reassure users and affected individuals that their rights are being actively supported. Ultimately, AI systems should deliver value to organisations by serving customers and supporting employees.

This is particularly important for AI systems that interact with or make decisions related to vulnerable and marginalised communities.

Conclusion: shaping the future of AI governance in Australia

Thoughtful and effective AI regulation will be central to Australia's ability to benefit sustainably from AI systems. Corporate leaders can support such efforts by demonstrating the innovative ways in which they deploy and govern fair, fit-for-purpose, accurate and accountable AI systems.

Achieving this will require investment by corporate leaders in their own skills and the capacity of their organisations to strategically and responsibly deploy AI systems.

This will often mean looking beyond the boundaries of their organisations. Corporate leaders should engage constructively with relevant industry, professional and standards associations to support the effective governance of AI across their industries. Corporate leaders should also engage with impacted communities and other stakeholders to ensure that AI systems are trustworthy, appropriately deployed and meet community expectations.

The next phase of HTI's AI Corporate Governance Program will be to develop insights from comparative governance frameworks, such as work health and safety, sustainability, and modern slavery, exploring their applicability to AI technologies and the governance challenges outlined in this report. Working with HTI's partners and stakeholders, we seek to identify a subset of systems, practices and capabilities that promise to support organisations to innovate responsibly and effectively with AI.



Table 12 – Questions for company directors and senior executives to support uplifting AI governance

	Questions for company directors	Questions for senior executives
Action 1: Build capacity and develop strategic expertise in AI	<ul style="list-style-type: none"> • Do all directors understand how, where, and why AI is being used in their organisation? • Do all directors appreciate their obligations under s180 regarding AI? • Is the board confident that executives possess the strategic expertise to execute to the strategy? • What additional support, advice or training do board members need to execute their duties with regard to using AI? • Do we have sufficient diversity in expertise and experience, or do we need to draw on outside capabilities? 	<ul style="list-style-type: none"> • Do all members of the executive team understand how, where, and why AI is being used in their organisation? • Do all members of the executive team possess the strategic expertise in AI to execute the strategy, realise opportunities, mitigate harms, and manage risks? • What additional support and strategic and technical expertise do internal operational teams – in particular, IT, data and analytics, legal, procurement, compliance, HR, and ESG – need to make effective decisions and deliver to the AI strategy? What external support is needed? • Do our front-line team members who rely on AI systems, or who deal with customers exposed to AI systems, possess the skills and information to faithfully execute their roles and identify potential failures as they emerge?
Action 2: Create a fit-for-purpose AI strategy	<ul style="list-style-type: none"> • Do we have a comprehensive AI strategy? • Does the AI strategy accurately leverage strategic foresight and reflect the evolving technical, commercial, regulatory, and social environment? • What external stakeholder perspectives should be brought to the board to strengthen our strategy? • Does the strategy identify at a high level both the business case for investing in AI, and the key risks and legal obligations relevant to the organisation? • How does or might our AI strategy intersect with other organisational strategies, particularly cyber security and data management? • Does the strategy clearly set a risk appetite for investment in AI systems, and is this in line with the board's expectations? 	<ul style="list-style-type: none"> • Do we have a comprehensive AI strategy? • Does the AI strategy accurately leverage strategic foresight and reflect the evolving technical, commercial, regulatory, and social environment? • What external stakeholder perspectives should be explored and deeply understood by the organisation to strengthen our strategy? • Are all relevant executives aware of the potential risks, opportunities and legal obligations posed by the use cases outlined in the strategy? • How does or might our AI strategy intersect with other organisational strategies, particularly cyber security and data management? • Which senior leaders bear responsibility for the strategy within the organisation, and how are they held to account?

Questions for company directors

Questions for senior executives

Action 3: Implement an integrated, comprehensive AI governance system

- Does the organisation have appropriate structures in place to support strategic discussion and effective decision making related to AI?
- What governance structures are in place to manage the operations of AI systems across the organisation?
- Are we confident that the organisation is effectively identifying, mitigating, and documenting key AI risks and opportunities?
- Are there effective mechanisms to ensure the board is appropriately and expeditiously informed around critical risks and emerging strategic opportunities related to AI systems?

- What is our current model of AI governance? How are AI systems identified, tracked and managed?
- What mix of governance models and processes are required for the organisation’s particular portfolio of AI systems?
- How is accountability distributed to ensure that appropriate individuals are responsible for governing and managing AI systems?
- What policies and practices are in place to identify, document, track and mitigate AI risks?
- What are the processes to keep senior management and the board informed of critical risks, governance failures and new opportunities?
- How can AI governance approaches integrate with existing impact assessment and risk management processes to be as efficient as possible?

Action 4: Set a human-centred AI culture

- Does the board have a unified view on how organisational values translate into its use of AI across functions and contexts?
- How are the views of stakeholders such as employees, customers and marginalised communities reflected in board discussion?
- Does the board appreciate the current culture of the organisation, and the level of alignment between stated mission and values and the embedded assumptions that drive behaviour?
- What actions can the board and directors take to model the desired culture?

- What do interactions and decisions by executives reveal about the current set of norms around AI?
- How do organisational goals, structure, and incentives influence the decisions related to AI? In what ways do these differ across functional teams and management levels?
- How can pre-existing values and ways of working core to organisational culture be extended to the use and development of AI?
- What structures and engagement supports the voice of impacted communities being integrated in decision-making processes?

HTI’s AI Corporate Governance Program takes a multi stakeholder approach to broaden the lens of corporate accountability in AI use in Australia today. We aim to support the transformation of AI governance systems and capabilities internal to organisations, as well as external policy and regulatory settings. For more information or to join the AI Governance Network, please contact hti@uts.edu.au.

Appendix: Common principles of responsible or ethical AI

	OECD values-based AI principles	US AI Bill of Rights	Canada: Our guiding principles	Singapore: Guiding principles	Australia's AI ethics principles
Safety	✓ Robustness, security and safety	✓ Safe and effective systems	✓ Protect system integration, national security		✓ Reliability & safety
Transparency	✓ Transparency & explainability	✓ Notice & explanation	✓ Transparent use & benefit.	✓ Transparent & explainable decision-making	✓ Transparency & explainability
Explainability	✓ Transparency & explainability	✓ Notice & explanation	✓ Provide meaningful explanations	✓ Explainable decision-making process	✓ Transparency & explainability
Fairness and quality	✓ Human-centred values and fairness	✓ Algorithmic discrimination protection		✓ Fair decision- making process	✓ Fairness
Accountability and oversight	✓ Accountability			✓ Accountability	✓ Accountability
Dispute resolution	✓ Human intervention & oversight	✓ Human alternatives & consideration	✓ Opportunities to challenge AI decisions		✓ Contestability
Right to object		✓ Human alternatives & consideration			✓ Contestability
Privacy		✓ Data privacy	✓ Protect personal information		✓ Privacy protection & security
Security	✓ Robustness, safety & security		✓ Protect system integration, national security		✓ Privacy, protection & security

Endnotes

1. PwC, *Sizing the prize: PwC's Global Artificial Intelligence Study: Exploiting the AI Revolution* (Report, 2017).
2. Eirini Kalliamvakou, 'Research: quantifying GitHub Copilot's impact on developer productivity and happiness' *Git Hub* (Blog Post, 7 September 2022) <<https://github.blog/2022-09-07-research-quantifying-github-copilots-impact-on-developer-productivity-and-happiness/>>.
3. Erik Brynjolfsson, Danielle Li and Lindsey Raymond, 'Generative AI at work' (Working Paper No 31161, National Bureau of Economic Research, 25 April 2023).
4. McKinsey, *The state of AI in 2022 – and a half decade in review* (Report, December 2022).
5. Reasons for this complexity include the rapid advancement and evolution of new methods (such as generative AI systems), varying contexts for use of the definition (such as research, or regulation and enforcement), and whether the audience is technical or non-technical. Defining AI from a technological perspective carries particular risk. If AI is defined too narrowly, then important technologies may be missed. If AI is defined too broadly, it may capture inappropriate systems and create unnecessary governance costs.
6. See, for example, Iqbal H. Sarker, 'AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems' (2022) 3(2) *SN Computer Science* 158.
7. Kate Crawford, *The Atlas of AI* (Yale University Press, 2021).
8. The bar of what is considered true 'artificial intelligence' has shifted rapidly from identifying handwriting, to beating humans at chess, to recognising human speech and beyond.
9. OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, 7; European Commission, *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts* COM (2021) 206 2021/0106/COD, art. 3(1) ("EU draft AI Act").
10. International standards on AI, such as ISO 22989, provide similarly broad definitions of AI systems. These would, for example, include the system involved in the Robodebt scandal, despite the fact that it deployed relatively unsophisticated algorithms.
11. National Institute of Standards and Technology (NIST), *Artificial Intelligence Risk Management Framework* (AI RMF 1.0) (NIST AI 100-1, January 2023).
12. 268 corporate leaders and strategic business decision makers were surveyed by HTI in the period from December 2022 to April 2023.
13. Austrade, *The 2021 Australian Artificial Intelligence export survey* (Report, 2021) 5.
14. See, for example, McKinsey, *The state of AI in 2022 – and a half decade in review* (Report, December 2022).
15. Committee for Economic Development of Australia (CEDA), *Artificial Intelligence: Principles to Practice* (Report, 2022).
16. Nestor Maslej, Loredana Fattorini, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Helen Ngo, Juan Carlos Niebles, Vanessa Parli, Yoav Shoham, Russell Wald, Jack Clark, and Raymond Perrault, *The AI Index 2023 Annual Report* (Stanford Institute for Human-Centered Artificial Intelligence Report, April 2023).
17. When prompted with specific examples, survey respondents and interviewees were often able to identify multiple, additional AI systems embedded within third party services.
18. One online study found that 43% of professional respondents said that they were using ChatGPT to complete work tasks, with two thirds of these doing so without their boss's knowledge: FishBowl, '70 percent of workers using ChatGPT at work are not telling their boss' (Blog Post, 2023) <<https://www.fishbowlapp.com/insights/70-percent-of-workers-using-chatgpt-at-work-are-not-telling-their-boss/>>.
19. Margot E Kaminski, 'Regulating the Risks of AI' (2023) 103 *Boston University Law Review* 7-8 (forthcoming).
20. Margot E Kaminski, 'Regulating the Risks of AI' (2023) 103 *Boston University Law Review* 7-8 (forthcoming), 7.
21. See, for example, Safiya Umoja Noble, *Algorithms of Oppression* (NYU Press, 2018); Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (St Martin's Press, 2018); Ruha Benjamin, *Race After Technology* (Polity, 2019); Patrick Williams and Eric Kind, *Data-driven policing: the hardwiring of discriminatory policing practices across Europe* (Report, November 2019); and Australian Human Rights Commission, *Human Rights and Technology Final Report* (Report, March 2021).
22. Lauren Smiley, 'I'm the Operator': The Aftermath of a Self-Driving Tragedy', *Wired* (online, 8 March 2022) <<https://www.wired.com/story/uber-self-driving-car-fatal-crash/>>.
23. Jonathan Turley, 'ChatGPT falsely accused me of sexually harassing my students. Can we really trust AI?', *USA Today* (online, 3 April 2023) <<https://www.usatoday.com/story/opinion/columnist/2023/04/03/chatgpt-misinformation-bias-flaws-ai-chatbot/11571830002/>>.
24. Dave Gershgorin, 'Black teen barred from skating rink by inaccurate facial recognition', *The Verge* (online, 16 July 2021) <<https://www.theverge.com/2021/7/15/22578801/black-teen-skating-rink-inaccurate-facial-recognition>>.
25. Joy Buolamwini and Timnit Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' (Conference Paper, Conference on Fairness, Accountability and Transparency PMLR 81, 2018) 77; K. S. Krishnapriya, Kushal Vangara, Michael C. King, Vitor Albiero and Kevin Bowyer, 'Characterizing the Variability in Face Recognition Accuracy Relative to Race' (Conference Paper, IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, June 2019); Inioluwa Deborah Raji and Joy Buolamwini, 'Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products,' (2023) 66(1) *Communications of the ACM* (Conference Paper, AAAI/ACM Conference on AI, Ethics, and Society, Association for Computing Machinery, January 2019) 101.
26. Josephine Wolff, *How to improve cybersecurity for artificial intelligence* (Brookings Institute Center for Technology Innovation Report, 9 June 2020).
27. *Australian Competition and Consumer Commission v Trivago NV* (2020) 142 ACSR 338 [13].
28. Consumer Policy Research Centre, *Duped by Design - Manipulative online design: Dark patterns in Australia* (Report, June 2022) 7.
29. Lily Hay Newman, 'AI Wrote Better Phishing Emails Than Humans in a Recent Test', *Wired* (online, 7 August 2021) <<https://www.wired.com/story/ai-phishing-emails/>>.
30. Office of the Australian Information Commissioner, 'OAIC opens investigations into Bunnings and Kmart' (Media Release, 12 July 2022) <<https://www.oaic.gov.au/newsroom/oaic-opens-investigations-into-bunnings-and-kmart>>.
31. Andrew White, 'Our Top Data and Analytics Predicts for 2019', *Gartner* (Blog Post, 3 January 2019) <https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/>.
32. See, for example, Charles Duhigg 'How Companies Learn Your Secrets' *New York Times* (online, 16 February 2012). <<https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1&r=1&hp>>.
33. Will Knight, 'The Apple Card Didn't 'See' Gender – and That's the Problem.' *Wired* (online, 19 November 2019) <<https://www.wired.com/story/the-apple-card-didnt-see-gender-and-thats-the-problem/>>.
34. Matthias Holweg, Rupert Younger, and Yuni Wen 'The Reputational Risks of AI' *California Management Review* (online, 24 January 2022) <<https://cmr.berkeley.edu/2022/01/the-reputational-risks-of-ai>>.
35. See, for example, Cait Kelly, 'Australian artists accuse popular AI imaging app of stealing content, call for stricter copyright laws', *The Guardian* (online, 12 December 2022) <<https://www.theguardian.com/australia-news/2022/dec/12/australian-artists-accuse-popular-ai-imaging-app-of-stealing-content-call-for-stricter-copyright-laws>>

36. See, for example, *Webb v Commonwealth Bank of Australia (Anti-Discrimination)* [2011] VCAT 1592.
37. We note the small sample size of these groups, reflecting indicative findings as further research is conducted.
38. See, generally, Ania Syrowatka, Masha Kuznetsova et al, 'Leveraging artificial intelligence for pandemic preparedness and response: a scoping review to identify key use cases' (2021) *npj Digit. Med.* 4, 96.
39. Bronwyn Hemsley, Emma Power et al, 'Will AI tech like ChatGPT improve inclusion for people with communication disability?' *The Conversation* (online, 19 January 2023) <<https://theconversation.com/will-ai-tech-like-chatgpt-improve-inclusion-for-people-with-communication-disability-196481>>.
40. See, for example, Bistra Dilkina, 'Artificial Intelligence and Conservation' (Seminar 1, Artificial Intelligence and Conservation Series, Fuller Science for Nature Fund, 11 October 2022).
41. Nathalie A Smuha, 'Beyond the individual: governing AI's societal harm' (2021) 10(3) *Internet Policy Review*.
42. Ada Lovelace Institute, *People, risk and the unique requirements of AI: 18 recommendations to strengthen the EU AI Act* (Policy Briefing, 31 March 2022) 10.
43. Emma Strubell, Ananya Ganesh and Andrew McCallum, 'Energy considerations for Deep Learning in NLP' (Conference Paper, Annual Meeting of the Association for Computational Linguistics, July 2019).
44. For a socio-technical perspective on how multiple social values influence the technology development process, see Thomas Philbeck, Nicholas Davis and Anne Marie Engtoft Larsen, *Values, Ethics and Innovation: Rethinking Technological Development in the Fourth Industrial Revolution* (World Economic Forum White Paper, August 2018).
45. Ziad Obermeyer et al, 'Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations' (2019) 366(6464) *Science* 447 <<https://www.science.org/doi/10.1126/science.aax2342>>
46. Josh Goldstein, Girish Sastry, Micah Musser, Renee DiResta, Matthew Gentzel, and Katerina Sedova, *Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations* (Georgetown University's Center for Security and Emerging Technology OpenAI and Stanford Internet Observatory Joint Report, January 2023).
47. Jane Wakefield, 'Deepfake presidents used in Russia-Ukraine war' BBC (online, 18 March 2022) <<https://www.bbc.com/news/technology-60780142>>.
48. See, for example, Zak Doffman, 'Hong Kong Exposes Both Sides of China's Relentless Facial Recognition Machine', *Forbes* (online, 26 August 2019) <<https://www.forbes.com/sites/zakdoffman/2019/08/26/hong-kong-exposes-both-sides-of-chinas-relentless-facial-recognition-machine/>>
49. Ada Lovelace Institute, *People, risk and the unique requirements of AI: 18 recommendations to strengthen the EU AI Act* (Policy Briefing, 31 March 2022) 10.
50. Australian Human Rights Commission, *Human Rights and Technology Final Report* (Report, March 2021) 77-83.
51. Pranshu Verma, 'The never-ending quest to predict crime using AI', *The Washington Post* (online, 15 July 2022) <<https://www.washingtonpost.com/technology/2022/07/15/predictive-policing-algorithms-fail/>>.
52. Rebecca Crotoof, *AI and the Actual IHL Accountability Gap* (Centre for International Governance Innovation Essay Series, 29 November 2022).
53. Maxim Naumov et al, 'Deep Learning Training in Facebook Data Centers: Design of Scale-up and Scale-out Systems' (arXiv, 18 August 2020) <<http://arxiv.org/abs/2003.09518>>
54. Jongsoo Park et al, 'Deep Learning Inference in Facebook Data Centers: Characterization, Performance Optimizations and Hardware Implications' (arXiv, 29 November 2018) <<http://arxiv.org/abs/1811.09886>>
55. David Patterson et al, 'The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink' (arXiv, 11 April 2022) <<http://arxiv.org/abs/2204.05149>>
56. Sasha Luccioni, 'The Mounting Human and Environmental Costs of Generative AI', *Ars Technica* (12 April 2023) <<https://arstechnica.com/gadgets/2023/04/generative-ai-is-cool-but-lets-not-forget-its-human-and-environmental-costs/>>
57. Pengfei Li et al, 'Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models' (arXiv, 6 April 2023) <<http://arxiv.org/abs/2304.03271>>
58. Nicole Gillespie, Steven Lockey, Caitlin Curtis, Javad Pool, & Ali Akbari, *Trust in Artificial Intelligence: A Global Study*. (The University of Queensland and KPMG Australia Report, 2023).
59. Steven Lockey, Nicole Gillespie, and Caitlin Curtis *Trust in Artificial Intelligence: Australian Insights* (The University of Queensland and KPMG Australia Report, October 2020).
60. International Data Corporation, 'Australia's Spending on Artificial Intelligence (AI) to Double to \$3.6 Billion by 2025, Says IDC' (Press Release, 19 May 2022).
61. Nicole Gillespie, Steven Lockey, Caitlin Curtis, Javad Pool, & Ali Akbari, *Trust in Artificial Intelligence: A Global Study*. (The University of Queensland and KPMG Australia Report, 2023).
62. Nicholas Davis, Lauren Perry and Edward Santow, 'Facial Recognition Technology: Towards a model law' (Human Technology Institute Report, September 2022).
63. By contrast, 81% of Chinese respondents, 71% of Brazilians, 69% of respondents in India and 59% of Singaporeans reported that they trust AI (Nicole Gillespie, Steven Lockey, Caitlin Curtis, Javad Pool, & Ali Akbari, *Trust in Artificial Intelligence: A Global Study*. (The University of Queensland and KPMG Australia Report, 2023)).
64. Nicole Gillespie, Steven Lockey, Caitlin Curtis, Javad Pool, & Ali Akbari, *Trust in Artificial Intelligence: A Global Study*. (The University of Queensland and KPMG Australia Report, 2023).
65. Nicole Gillespie, Steven Lockey, Caitlin Curtis, Javad Pool, & Ali Akbari, *Trust in Artificial Intelligence: A Global Study*. (The University of Queensland and KPMG Australia Report, 2023).
66. Generative AI systems include so-called 'large language models' or LLMs (such as Google's BERT, Bard and LaMDA, OpenAI's ChatGPT series and BigScience's BLOOM), latent diffusion-based image creators (such as Google's DALL-E, Midjourney, and Stability AI's Stable Diffusion). Other examples of LGAIMs include AI-based music generators (such as AIVA, Ecrett Music and OpenAI's MuseNet), and AI video generators (such as Synthesia and Meta's Make-A-Video).
67. See, for example, the decision of the Italian Data Protection Authority to impose a temporary ban on the processing of Italian user's data by ChatGPT in March 2023 (GPDP, 'Intelligenza artificiale: il Garante blocca ChatGPT. Raccolta illecita di dati personali. Assenza di sistemi per la verifica dell'età dei minori' (Press Release, 31 March 2023) <<https://www.gpdp.it/home/docweb/-/docweb-display/docweb/9870847>>).
68. This is a known problem for generative AI. To counter this, OpenAI's 'InstructGPT' model used reinforcement learning from human feedback to fine-tune models to reduce toxicity and 'hallucinations', while increasing truthfulness and appropriateness. Despite this, OpenAI stated that "our InstructGPT models are far from fully aligned or fully safe; they still generate toxic or biased outputs, make up facts, and generate sexual and violent content without explicit prompting."
69. See, for example, Prarthana Prakash, 'ChatGPT falsely accused a mayor of bribery when he was actually the whistleblower—now he wants to sue in what could be the first defamation case against a bot', *Fortune* (online, 6 April 2023) <<https://fortune.com/2023/04/05/chatgpt-falsely-accused-australian-mayor-bribery-openai-defamation/>>.
70. Linda J. Skitka, Kathleen Mosier, Mark D. Burdick 'Accountability and automation bias' (2000) 52(4) *International Journal of Human-Computer Studies* 701.
71. Emily M Bender, Timnit Gebru, Angelina McMillan-Major, Shmargaret Shmitchell 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?' (Conference Paper, ACM Conference on Fairness, Accountability, and Transparency, March 2021).
72. Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, William Fedus, 'Emergent Abilities of Large Language Models' (2022) *Transactions on Machine Learning Research*.

73. Nicholas Davis, Lauren Perry and Edward Santow, 'Facial Recognition Technology: Towards a model law' (Human Technology Institute Report, September 2022).
74. Phillip Hacker, Andreas Engel and Marco Mauer, 'Regulating ChatGPT and other Large Generative AI Models' (Working Paper, 5 April 2023).
75. Jo Constance, 'Nearly a Third of White-Collar Workers Have Tried ChatGPT or Other AI Programs, According to a New Survey' *Time* (online, 19 January 2023).
76. Justice N.J. Owen, *The Failure of HIH Insurance: A corporate collapse and its lessons* (Report, April 2003), xxxiv.
77. Australian Institute of Company Directors (AICD), 'What is Good Governance?' *Good governance* <<https://www.aicd.com.au/good-governance.html>>.
78. Fifth quadrant, Ethical AI Advisory and Gradient Institute '2021 Responsible AI Index' (Report, 2021).
79. McKinsey, *The state of AI in 2022—and a half decade in review* (Report, December 2022).
80. For example, Microsoft AI Principles (2018), AI at Google: Our Principles (2018), IBM Everyday Ethics for AI (2022). Several Australian businesses, including National Australia Bank, Commonwealth Bank, Microsoft and Telstra also participated in the Australian AI Ethics Principles pilot to test Australia's Artificial Intelligence Ethics Framework.
81. Matthew Beard, and Simon Longstaff, 'Ethical By Design: Principles for Good Technology' (The Ethics Centre Report, 2018).
82. Thilo Hagendorff, 'The Ethics of AI Ethics: An Evaluation of Guidelines' (2020) 30 *Minds and Machines* 99.
83. Luke Munn, 'The uselessness of AI ethics' (2022) *AI and Ethics*.
84. Thilo Hagendorff, 'AI ethics and its pitfalls: not living up to its own standards?' (2023) 3 *AI and Ethics* 329.
85. Safe Work Australia, 'The Effectiveness Of Work Health And Safety Interventions By Regulators: A Literature Review' (Report, April 2013).
86. Luciano Floridi, 'Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical' (2019) 32 *Philosophy & Technology* 185.
87. Ellen Sheng, 'In Generative AI Legal Wild West, the Courtroom Battles Are Just Getting Started', CNBC (3 April 2023) <<https://www.cnbc.com/2023/04/03/in-generative-ai-legal-wild-west-lawsuits-are-just-getting-started.html>> <https://www.cnbc.com/2023/04/03/in-generative-ai-legal-wild-west-lawsuits-are-just-getting-started.html>
88. Organisations operating in Australia may also be subject to foreign legislation. For example, in certain circumstances, the EU GDPR applies to Australian businesses that deal with the data of EU individuals. However, this report only examines relevant Australian legislation.
89. This table does not cover all possible harms and legal obligations as a number of obligations are industry or context-specific.
90. See, for example, Bret Walker and Gerald NG, *Australian Institute of Company Directors The Content of Directors' "Best Interest" Duty* (Memorandum of Advice, 24 February 2022).
91. *Corporations Act 2001* (Cth), s 180.
92. Rosemary Teele Langford and Andrew Godwin, *Directors' duties and cyber security - it's complicated* (Pursuit Report, 3 September 2021).
93. See, for example, Australian Securities and Investment Commission, *Cyber resilience: Health Check* (Report 429, March 2015), 43; AICD, *Cyber Security Governance Principles* (Report, October 2022).
94. See, for example, *Cassimatis v Australian Securities and Investment Commission* [2020] 275 FCR 533.
95. *Cassimatis v Australian Securities and Investment Commission* (2016) 336 ALR 209 at 301 [482].
96. Bret Walker and Gerald NG, *Australian Institute of Company Directors The Content of Directors' "Best Interest" Duty* (Memorandum of Advice, 24 February 2022).
97. This list is not exhaustive and additional laws will apply, such as surveillance laws, and laws specific to the relevant sector or use case of the technology.
98. The Privacy Act does not apply to small businesses with a turnover of less than \$3 million or specific organisation dealing in specific types of personal information such as health services providers and credit reporting bodies.
99. Whilst the possibility of a common law tort of privacy was recognised in *Australian Broadcasting Corporation Lenah Game Meats* (2001) 185 ALR 1, the courts have since interpreted that decision narrowly. The existence of this tort has not been confirmed by an Australian appellate court.
100. The Australian Government is currently considering a proposal to introduce a statutory tort of privacy for serious invasions of privacy (Australian Government, *Privacy Act Review* (Report, 2022) 15, Proposal 27.1).
101. This issue was raised in *Commissioner initiated investigation into Clearview AI, Inc. (Privacy)* [2021] AICmr 54 (14 October 2021).
102. Australian Government, *Privacy Act Review* (Report, 2022).
103. Under Australian Consumer Law, consumers relevantly include a person or business who acquires goods and services for less than \$100,000 or which are ordinarily acquired for personal, domestic or household use or consumption.
104. *Corporations Act 2001* (Cth), the *National Consumer Credit Protection Act 2009* (Cth), *National Energy Retail Law* (South Australia) 2011 (Cth), *National Energy Retail Rules* (Cth), *National Energy Retail Regulations* (Cth) and *Telecommunications Consumer Protection Code* (Cth).
105. *Corporations Act 2001* (Cth), s 912A.
106. *Australian Securities and Investments Commission v RI Advice Group Pty Ltd* [2022] FCA 496.
107. ASX, *Listing Rules* (at 1 December 2019), Rule 3.1.
108. An eligible data breach occurs there is unauthorised access to or unauthorised disclosure of personal information, or a loss of personal information, that an organisation or agency holds this is likely to result in serious harm to one or more individuals, and the organisation or agency hasn't been able to prevent the likely risk of serious harm with remedial action (Privacy Act, s 26WE(2)).
109. The Privacy Act Review Report proposes making the deadline for reporting eligible data breaches to the OAIC to 72 hours, with affected individuals to be notified as soon as practicable (Australian Government, *Privacy Act Review* (Report, 2022), 15, Proposal 28.2). Under the current regime, where an entity has reasonable grounds to suspect, but does not yet believe, that an eligible data breach has occurred, it has a 30-day period to complete its assessment of the breach.
110. Australian Human Rights Commission, *Using artificial intelligence to make decisions: Addressing the problem of algorithmic bias* (Report, 2020).
111. *Racial Discrimination Act 1975* (Cth); *Sex Discrimination Act 1984* (Cth); *Disability Discrimination Act 1992* (Cth); *Age Discrimination Act 2004* (Cth); *Discrimination Act 1991* (ACT); *Anti-Discrimination Act 1977* (NSW); *Anti-Discrimination Act* (NSW); *Anti-Discrimination Act 1992* (NT); *Anti-Discrimination Act 1991* (Qld); *Equal Opportunity Act 1984* (SA); *Anti-Discrimination Act 1998* (Tas); *Equal Opportunity Act 2010* (Vic); and *Equal Opportunity Act 1984* (WA).
112. Law of negligence established in case law including: *Donoghue v Stevenson* [1932] UKHL 100; *Todman v Victa Ltd* [1982] VR 849; *Tabet v Gett* (2010) 240 CLR 537 (among others). In some jurisdictions and in certain circumstances, the law of negligence has been codified e.g. *Civil Liability Act 2002* (NSW), *Civil Liability Act 2003* (Qld), *Wrongs Act 1958* (Vic), *Civil Liability Act 1936* (SA), *Civil Liability Act 2002* (WA).
113. Various federal, state and territory laws, including *Work Health and Safety Act 2011* (Cth); *Work Health and Safety Act 2011* (ACT); *Work Health and Safety Act 2011* (NSW); *Work Health and Safety (National Uniform Legislation) Act 2011* (NT); *Work Health and Safety Act 2011* (Qld); *Work Health and Safety Act 2012* (SA); *Work Health and Safety Act 2012* (Tas); *Occupational Health and Safety Act 2004* (Vic); *Occupational Safety and Health Act 1984* (WA) (among others).
114. *Ministerial Declaration The G7 Digital and Tech Ministers' Meeting* (30 April 2023).
115. US Chamber of Commerce, *Commission on Artificial Intelligence Competitiveness, Inclusion, and Innovation Report and Recommendations* (Report, 2023).
116. Brad Smith, 'Meeting the AI moment: advancing the future through responsible AI' *Microsoft* (Blog Post, 2 February 2023).

117. Sundar Pichai, 'Why Google thinks we need to regulate AI' *Financial Times* (online, 19 January 2020) <<https://www.ft.com/content/3467659a-386d-11ea-ac3c-f68c10993b04>>.
118. Kevin Stankiewicz, 'IBM chief calls for 'precision regulation' on AI that weighs privacy against benefits to society' CNBC (online, 22 January 2020) <<https://www.cnbc.com/2020/01/22/ibm-ceo-ginni-rometty-calls-for-precision-regulation-on-ai.html>>.
119. Cecilia Kang, 'OpenAI's Sam Altman Urges A.I. Regulation in Senate Hearing', *The New York Times* (online, 16 May 2023) <<https://www.nytimes.com/2023/05/16/technology/openai-altman-artificial-intelligence-regulation.html>>
120. Amelia Heathman, 'Salesforce CEO Marc Benioff at Davos: the tech industry needs regulating' *Verdict* (Blog Post, 23 January 2018) <<https://www.verdict.co.uk/salesforce-marc-benioff-davos/>>
121. The OECD AI Policy Observatory's repository of AI policy instruments includes those relating to guidance and regulation (i.e. regulatory interventions), governance (i.e. regulatory initiatives, public consultation, AI use in the public sector), financial support (i.e. research grants and funding), and AI enablers and other incentives (i.e. skills and education, data sharing, AI infrastructure, networking, awards).
122. AI-related bills that were passed into law identified by searches of the keywords "artificial intelligence" on the legislative records of 127 countries.
123. HTI analysis based on data from OECD.AI, 'National AI policies & strategies' OECD. *AI Policy Observatory* (Database, 2023); Nestor Maslej, Loredana Fattorini, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Helen Ngo, Juan Carlos Niebles, Vanessa Paril, Yoav Shoham, Russell Wald, Jack Clark, and Raymond Perrault, *The AI Index 2023 Annual Report* (Stanford HAI, 2023)(Stanford Institute for Human-Centered Artificial Intelligence Report, April 2023).
124. OECD.AI, 'National AI policies & strategies' OECD. *AI Policy Observatory* (Database, 2023).
125. Parliament, European, 'Proposal for a Regulation on a European Approach for Artificial Intelligence | Legislative Train Schedule', European Parliament <<https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence>>
126. 'AI Act: A Step Closer to the First Rules on Artificial Intelligence | News | European Parliament' (5 November 2023) <<https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>>
127. See for example analysis in Alex Engler, 'The EU AI Act Will Have Global Impact, but a Limited Brussels Effect', Brookings (8 June 2022) <<https://www.brookings.edu/research/the-eu-ai-act-will-have-global-impact-but-a-limited-brussels-effect/>>
128. HTI analysis based on data from OECD.AI, 'National AI policies & strategies' OECD. *AI Policy Observatory* (Database, 2023); Nestor Maslej, Loredana Fattorini, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Helen Ngo, Juan Carlos Niebles, Vanessa Paril, Yoav Shoham, Russell Wald, Jack Clark, and Raymond Perrault, *The AI Index 2023 Annual Report* (Stanford HAI, 2023)(Stanford Institute for Human-Centered Artificial Intelligence Report, April 2023).
129. Nestor Maslej, Loredana Fattorini, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Helen Ngo, Juan Carlos Niebles, Vanessa Paril, Yoav Shoham, Russell Wald, Jack Clark, and Raymond Perrault, *The AI Index 2023 Annual Report* (Stanford HAI, 2023)(Stanford Institute for Human-Centered Artificial Intelligence Report, April 2023).
130. HTI analysis based on data from OECD.AI, 'National AI policies & strategies' OECD. *AI Policy Observatory* (Database, 2023).
131. HTI analysis based on data from OECD.AI, 'National AI policies & strategies' OECD. *AI Policy Observatory* (Database, 2023); Nestor Maslej, Loredana Fattorini, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Helen Ngo, Juan Carlos Niebles, Vanessa Paril, Yoav Shoham, Russell Wald, Jack Clark, and Raymond Perrault, *The AI Index 2023 Annual Report* (Stanford HAI, 2023)(Stanford Institute for Human-Centered Artificial Intelligence Report, April 2023).
132. Charlotte Siegmund and Markus Anderjurg, 'The Brussels Effect and Artificial Intelligence: How EU Regulation Will Impact the Global AI Market', arXiv.org (23 August 2022) <<https://arxiv.org/abs/2208.12645v1>>
133. Casey Tonkin, 'Government prepared to step in with AI regulation' ACS (ICT News, 23 March 2023) <<https://ia.acs.org.au/article/2023/govt-prepared-to-step-in-with-ai-regulation.html>>.
134. Australian Government Department of Industry, Science and Resources, *Australia's AI Ethics Principles* (2019).
135. Standards Australia, Introduction to Standards for Artificial Intelligence (Report, May 2023)
136. Peter Cihon, 'AI & Global Governance: Using International Standards as an Agile Tool for Governance' *United Nations University Centre for Policy Research* (Article, 8 July 2018) <<https://cpr.unu.edu/publications/articles/ai-international-standards.html>>
137. European Parliament, *Draft Compromise Amendments on the Draft Report: Proposal for a regulation of the European Parliament of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act)*
138. Hadrien Pouget, 'Standard Setting' *The AI Act* (Web Page) <<https://artificialintelligenceact.eu/standard-setting/>>
139. Ada Lovelace Institute, *Inclusive AI governance: civil society participation in standards development* (Discussion Paper, March, 2023).
140. NIST, *Artificial Intelligence Risk Management Framework* (AI RMF 1.0) (NIST AI 100-1, January 2023).
141. 'P2863 Recommended Practice for Organizational Governance of Artificial Intelligence', *IEEE Standards Association* (Web Page) <<https://standards.ieee.org/ieee/2863/10142/>>.
142. Thomas Philbeck, Nicholas Davis and Anne Marie Engtoft Larsen, *Values, Ethics and Innovation: Rethinking Technological Development in the Fourth Industrial Revolution* (World Economic Forum White Paper, August 2018).
143. OECD.AI, 'OECD AI Principles overview' OECD. *AI Policy Observatory* (Web Page, May 2019) <<https://oecd.ai/en/ai-principles>>.
144. The White House, *Blueprint for an AI Bill of Rights Making Automated Systems Work for the American People* (Blueprint, October 2022).
145. Government of Canada, 'Our guiding principles' *Responsible use of artificial intelligence* (Web page, 25 April 2023) <<https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html>>.
146. Personal Data Protection Commission, *Model Artificial Intelligence Governance Framework Second Edition* (Model Framework 2nd Edition, 21 January 2020).
147. Australian Government Department of Industry, Science and Resources, *Australia's AI Ethics Principles* (2019).



For more information

Human Technology Institute
hti@uts.edu.au

University of Technology Sydney
PO Box 123
Broadway NSW 2007

uts.edu.au