



ARTIFICIAL INTELLIGENCE AND HUMAN RIGHTS

INVESTOR TOOLKIT

Navigating the nexus of AI and
human rights: A toolkit for investors
in a world of rapid change

Disclaimer

The views expressed in this toolkit are those of the authors and not those of their organisations, nor that of RIAA. Investors, or any other party referring to, or using this toolkit in part or whole, should undertake their own research before commencing engagement with investee companies. All care has been taken in the preparation of this toolkit, but the contents of it cannot be relied upon.

Contact us

RESPONSIBLE INVESTMENT ASSOCIATION AUSTRALASIA

**Level 2, 696 Bourke Street
Melbourne, VIC 3000
Australia**

**Tel: +61 3 7068 9456
info@responsibleinvestment.org
responsibleinvestment.org**

ACKNOWLEDGEMENT

This toolkit was prepared by RIAA's Human Rights Working Group (Digital Tech subgroup), which is a collective of representatives from the Australian and Aotearoa New Zealand investment community.

The following people contributed to the toolkit:

- **Mark Lyster** (co-chair of Digital Tech Subgroup) – Lyster & Associates/Edge Impact
- **Liza McDonald** (co-chair of Digital Tech Subgroup) – Head of Responsible Investment, Aware Super
- **Belinda White** (Editor)
- **Janelle Morrison** – ESG Analyst, Ausbil Investment Management Limited
- **Jessica Wyndham** – Associate Director, KPMG Banarra
- **Moana Nottage** – ESG and Sustainability Analyst, Alphinity Investment Management
- **Jessica Cairns** – Head of ESG & Sustainability, Alphinity Investment Management
- **Amy Krizanovic** – Head of ESG, Magellan Asset Management
- **Jacqueline Stokes** – ESG Analyst, Magellan Asset Management
- **Emma Pringle** – Head of ESG and Portfolio Manager, Maple-Brown Abbott
- **Nick Dexter** – Principal Consultant, Edge Impact
- **Kaitlin Priestley** – ESG & Sustainability Manager, Pinnacle Investment Management Group
- **Yajaswi Rai** – Masters in Sustainability from The University of Sydney and Intern at the Responsible Investment Association Australasia
- **Alison Ewings** – General Manager ESG, QIC
- **Tyrone Louw** – Manager, Responsible Investments – Research, Aware Super
- **Måns Carlsson, OAM** (Chair, RIAA Human Rights Working Group) – Head of ESG Research, Ausbil Investment Management Limited

Thanks to the following external reviewers who provided valuable feedback:

- **Investor Alliance for Human Rights:**
Anita Dorett (Director)
- **Australian eSafety Commissioner (eSafety)**

Oversight and project management support provided by:

- **Responsible Investment Association Australasia:**
Estelle Parker, co-CEO; Isabella Marotta, Program Officer - Working Groups; Jessica Rowe, Manager of Working Groups; Antonia Bowlen, Project Officer - Working Groups; and Ada Tso, Manager of Communications and Media.

CONTENTS

ACKNOWLEDGEMENT	3	FIGURES	
OVERVIEW	5	FIGURE 1 AI archetypes arranged by operational and autonomous nature	9
GLOSSARY	6	FIGURE 2 Australia's AI Ethics Principles	10
INTRODUCTION	7	FIGURE 3 UNPRI human rights due diligence process	15
SECTION 1: ARTIFICIAL INTELLIGENCE OVERVIEW	8	FIGURE 4 Simplified overview of the AI supply chain from the OCED	17
1.1 Overview of AI	8	FIGURE 5 OECD AI Classification Framework	18
1.2 Salient human rights risks and issues	8	FIGURE 6 The 5D Framework	19
1.3 Emergence of regulatory, governance and ethical frameworks	10	FIGURE 7 Monash University and OHCHR Degree of risk to business	20
SECTION 2: WHY SHOULD INVESTORS CARE ABOUT THE HUMAN RIGHTS IMPACTS OF ARTIFICIAL INTELLIGENCE	12	GRAPH 1 Social shareholder proposals US technology sector	25
2.1 Reputational and operational risk	12	GRAPH 2 Human rights related shareholder proposals US technology sector	25
2.2 Regulatory risk	12	TABLE 1 Type of AI systems	8
2.3 Financial risk	13	TABLE 2 Overview of Artificial intelligence risks: Categories and examples of harm from AI systems (adapted from HTI's categorisation of risks)	14
2.4 Risks of harm to people	13		
SECTION 3: INTEGRATION – ASSESSING AI-RELATED HUMAN RIGHTS IMPACTS	15		
3.1 Introduction	15		
3.2 Underlying concepts	16		
3.3 Framework for identifying AI-related human rights risks/issues	16		
3.4 Identifying and assessing human rights impacts	19		
3.5 Identifying the salient human right risks/issues	19		
3.6 Assessing risk mitigation	20		
3.7 Good AI governance practice	20		
3.8 Assessing the maturity of human rights risk management	21		
3.9 Quantification of potential AI risks (financial and non-financial)	22		
Section 3 framework graphic	23		
SECTION 4: STEWARDSHIP – HOW INVESTORS CAN ENGAGE COMPANIES ON AI RISKS	24		
4.1 Prioritisation	24		
4.2 Engagement approaches	24		
ROAD MAP	27		
APPENDICES	28		
Appendix A: Generic AI-related human rights due diligence and stewardship guide	28		
Appendix B: AI-related human rights risk matrix	30		
Appendix C: Specific human rights engagement guide	32		
Appendix D: Further resources	33		
Endnotes	34		

OVERVIEW

The rapid growth of Artificial Intelligence (AI) presents both opportunities and challenges for investors, especially regarding human rights impacts (referred to as 'AI-related human rights risks'). This toolkit is designed to help investors understand and navigate the key challenges, risks and human rights risks/issues in relation to AI.

The development of the toolkit has been a collaborative effort of RIAA's Human Rights Working Group (Digital Tech Subgroup) and includes contributions from investment and human rights professionals across Australia and Aotearoa New Zealand. It offers practical guidance for assessing risks, engaging with companies and advocating for policy change.

The primary goal of this toolkit is to enhance the investment community's understanding of the financial and human rights risks associated with AI.

This toolkit is organised into four sections that progress from foundational knowledge to practical guidance for investment decision-making and engagement:

- **Introduction to AI and human rights:** Sets the stage for the critical nexus between AI and its societal implications, establishing the foundations for the toolkit's objectives.
- **Overview – Key human rights risks for AI:** Offers a deep dive into relevant AI technologies and their human rights impacts, backed by case studies illustrating the real-world consequences of the deployment of AI technology.
- **Integration – Assessing financial and salient human rights risks/issues of AI:** Focuses on methodologies for assessing the financial risks and salient human rights risks/issues of digital technologies, and provides tools for due diligence and risk management.
- **Stewardship – Engaging with companies on AI risks:** Details strategies for investor engagement, highlighting approaches to influence corporate practices within investee companies in investment portfolios and advocate for responsible AI technology use.

This toolkit aims to provide investors with the knowledge and tools to proactively promote responsible investment practices that protect, respect and remedy human rights impacts. It is acknowledged that there may be broader societal implications from the use of AI e.g. large scale transition of workforces and unemployment, however, this is beyond the scope of this toolkit.

This toolkit has been developed by the Responsible Investment Association Australasia's (RIAA) Human Rights Working Group (Digital Tech sub-group) and its members to:

- 1 increase members' **knowledge and awareness** of the risks and human rights impacts associated with AI (Sections 1 & 2);
- 2 provide members with **practical guidance to assess risks (including financial) and human rights risks/issues, and prioritise stewardship activities** as part of their investment decision-making (Sections 3 & 4); and
- 3 build members' confidence in **advocating for appropriate public policy and regulatory change** required to prevent systemic risks and protect the vulnerable.

While much has been written about the investor-relevant opportunities from AI, this toolkit predominantly focuses on the risks involved and how these can be avoided/mitigated/reduced. The information provided in this toolkit is not exhaustive and should be treated as a supplementary source of information to complement users' own investment and stewardship processes.

GLOSSARY

Algorithm. An algorithm is a step-by-step set of instructions, or a defined set of rules designed to perform a specific task, solve a particular problem, or achieve a desired outcome. In the context of AI, an algorithm can be used by machine learning systems to make predictions based on sets of training data.

Chatbot. A chat robot ('chatbot') is a computer program designed to simulate conversation with users, especially through text or voice interactions. Powered by artificial intelligence or predefined rules, chatbots engage in natural language conversations, providing information, answering queries, or assisting users with various tasks.

Deepfake. A deepfake is a manipulated or synthesised multimedia content, typically using deep learning and/or generative AI techniques, where AI is employed to replace, alter, or superimpose existing images or videos with highly realistic, but fabricated, content. Deepfakes often involve the use of sophisticated algorithms to create convincing simulations of real individuals, raising concerns about misinformation, identity theft and the potential for deceptive content in various media.

Generative AI. Generative AI refers to a class of AI models and algorithms that can generate new, original content. Unlike traditional AI systems that are trained to recognise patterns in existing data, generative AI can create new data that wasn't part of its training set, allowing for versatile applications including in various creative fields.

Machine learning. Machine learning is a subset of AI that focuses on the development of algorithms and statistical models that enable computers to improve their performance on a specific task over time without being explicitly programmed. The primary goal of machine learning is to enable computers to learn from data and make predictions or decisions based on that learning.

Social scoring. Refers to a system used to assess and rank individuals based on various social and behavioural criteria. These criteria can include factors such as financial behaviour, social interactions, online activity and adherence to societal norms or regulations. Social scoring systems can be used for reasons such as risk assessment and credit evaluation, as well as by social media platforms to measure and rank user influence, popularity, or engagement levels.

Training data. Training data refers to a set of examples or instances used to train a machine learning model. These data points are inputs into the model during the learning phase, enabling it to recognise patterns, correlations and relationships to make predictions or classifications.

INTRODUCTION

Society is living through a period of significant transformation, driven in part by exponential growth in the development and application of Artificial Intelligence (AI). The dramatic rise of AI over recent decades has been a catalyst for rapid and widespread change, creating both excitement and fear about its potentially transformative effects.

The potential benefits of AI for society and companies are immense, in areas including healthcare, access to information, economic modelling, disaster mitigation and infrastructure management. The benefits of AI to business can include greater efficiency, increased productivity, expanded markets and improved quality. As such, the estimated economic value of AI is considerable and can represent significant growth opportunities to companies, and therefore investors.

At the same time, there is growing awareness about the risks, some potentially catastrophic, posed by AI when inadequately designed, inappropriately or maliciously deployed or overused. A related issue is the inadequate consideration of adverse human rights impacts when new products and services using AI are deployed, often in a bid to gain the first mover's advantage in the market.

In human rights terms, AI poses risks of bias and exacerbates systemic discrimination of individuals, particularly those from historically marginalised communities. AI can increase system vulnerability to cyber-attacks resulting in privacy violations at scale. AI can facilitate targeted attacks on vulnerable groups, particularly children, in addition to other human rights abuses.

RIAA MEMBER CONCERNS

Investor concerns about digital technology impacts on human rights were polled at RIAA's 2023 annual conference in Australia. The top five issues of concern were:

- Privacy and data protection
- Political participation – disinformation, polarisation, barriers to democracy
- Online safety
- Discrimination
- Conflict and security

While the question was broader than AI, each issue can be linked to the use, and potential misuse, of the technology associated with AI.

Relevance to investors

From an investor perspective, businesses that have failed to put in place adequate governance safeguards around the design and deployment of AI are at risk of suffering financial consequences and some have been subject to sizeable fines. The result has been reputational damage and serious challenges to these businesses' social licence to operate and grow. In addition, the legal risks will likely increase as the regulatory framework continues to evolve to keep pace with the technology change.

The United Nations Guiding Principles on Business and Human Rights (UNGPs) establishes a framework of principles for governments and business enterprises to “protect, respect and remedy human rights impacts”¹. Institutional investors can support the implementation of the UNGPs in several ways, have a responsibility to respect human rights (as defined by the UNGPs) and have a vital role to play in ensuring business alignment with the UNGPs. This can be ensured through the application of two particular levers:

- **Integration** – ongoing consideration of the implications of AI within an investment analysis and decision-making process with the aim of improving risk-adjusted returns; undertaking due diligence to assess both the potential risks to people and to asset values from the design or deployment of AI within the investment decision-making processes.
- **Stewardship** – using investor ownership rights and influence to protect and enhance overall long-term value for clients and beneficiaries, by seeking to establish that investee companies have in place:
 - Appropriate governance, including effective stakeholder engagement processes, ongoing systems of impact monitoring and, fair and transparent remediation systems.
 - The proper use of AI and robust security of related data.
 - The due care and caution in the design and sale of AI systems for third party use.
 - Advocating for policy settings supportive of the appropriate management of risks related to the application of AI.

SECTION 1: ARTIFICIAL INTELLIGENCE OVERVIEW

1.1 Overview of AI

There is no universally applied definition of ‘artificial intelligence’. However, a helpful description of AI used by the Human Technology Institute (HTI) of the University of Technology Sydney draws from two globally recognised and broadly applied definitions of AI, one contained within the AI principles of the Organisation for Economic Co-operation and Development and the other the AI Act (2024) of the European Union. The description is as follows:

“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.”²

This description encompasses multiple sorts of AI systems, outputs and use cases. Two important subsets of AI are:

- **Machine learning systems** – trained on pre-existing, often unstructured, data and apply lessons from the past to new data, to make predictions for the future.
- **Expert systems** – that solve complex problems by applying ‘if-then’ and logical reasoning to a knowledge base to mimic human decision-making processes.

Table 1 provides provides examples of the ways in which different forms of AI are used in practice.

1.2 Salient human rights risks and issues

As discussed above, key investment risks include companies’ social licence to operate and grow, as well as potential legal risks. In considering these and other human rights-related investment risks associated with the use of AI, including future regulatory developments that may impact companies (and investors), a useful starting point is to draw on the existing internationally recognised human rights frameworks, as these may shape the future regulatory framework around AI globally, such as:

1. The International Bill of Rights (consisting of the Universal Declaration of Human Rights and two binding treaties to which Australia is a party; the International Covenant on Civil and Political Rights; and the International Covenant on Economic, Social and Cultural Rights); and
2. Seven other core international human rights treaties, each focused on a specific collection of rights (e.g. right to freedom from torture) or groups of people (e.g. rights of women, children and persons with disabilities).¹
3. The fundamental conventions of the International Labour Organization (ILO).³
4. In a business context, the UNGPs establishes a framework of principles for governments and business enterprises to “protect, respect and remedy human rights impacts”.

¹ The seven other core human rights treaties include: the Convention on the Elimination of All Forms of Discrimination against Women, the Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment, Convention on the Rights of Persons with Disabilities, the International Convention on the Elimination of All Forms of Racial Discrimination, the Convention on the Rights of the Child, the International Convention on the Protection of the Rights of All Migrant Workers and Members of their Families and the International Convention for the Protection of All Persons from Enforced Disappearance.

TABLE 1 Type of AI systems

Type of AI system	Example Use Cases
Natural language systems	Translation, autocorrect, conversational AI (e.g. Siri and Alexa), speech recognition
Generative AI	Text generation (e.g. ChatGPT, Gemini), image and video generation (e.g. Dall-E, Midjourney), chatbots and other virtual agents
Facial recognition technologies	One-to-one matching (e.g. to unlock a phone), one-to-many identification to check against a database to identify a person (e.g. in criminal investigations)
Recommender systems	Product recommendations (e.g. targeted advertisements), service or information recommendations (e.g. Spotify, Netflix)
Automated decision-making systems	E.g. credit determinations, home loan determinations, hiring decisions
Robotic process automation	E.g. website scraping including for research or comparative purposes

The potential rights harmed by AI are as vast as its applications, which in turn, can impact the value of investments in the previously mentioned use cases and other ways. Examples include:

Right to non-discrimination: One of the most common concerns about AI is the risk of discrimination, often caused by the reliance of AI systems on historical data that leads to the replication and amplification of existing patterns of exclusion and hence perpetuates systemic discrimination. When a biased outcome impacts an individual's ability to access, for example, education, employment, or housing, then additional human rights will be impacted. This may lead to various reputational and legal risks.

Right to mental and physical health: For example, generative AI is being used to target women and children, in particular, through the creation of sexually explicit deepfakes that violate their rights to non-discrimination and privacy and cause serious harm to their mental and physical health. In addition, AI supported algorithms that power many social media platforms can be found to amplify negative and destructive content. This may also lead to reputational and legal risks.

Right to privacy: AI systems, fed on enormous swathes of often highly personal and sensitive data, can be vulnerable to hacks, leading to violations of the right to privacy via the release of private personal information, including health, intellectual property and financial information. This may lead to reputational and legal risks, particularly as privacy regulation continues to tighten.

Right to liberty and security of person: The personal liberty of individuals can be violated when facial recognition technologies lead to arbitrary arrest and when AI algorithms are used to predict behaviour, including the likelihood that individuals applying for bail or parole will reoffend. This may lead to reputational and legal risks.

Right to remedy: Any individual who is harmed by an AI system or its output has the right to receive a remedy. Remediation is made difficult by the opacity of AI systems, their 'black box' nature and by a lack of transparency about when and how AI systems have been used. Naturally, this opens up a series of legal risks that can impact companies and, therefore, investments.

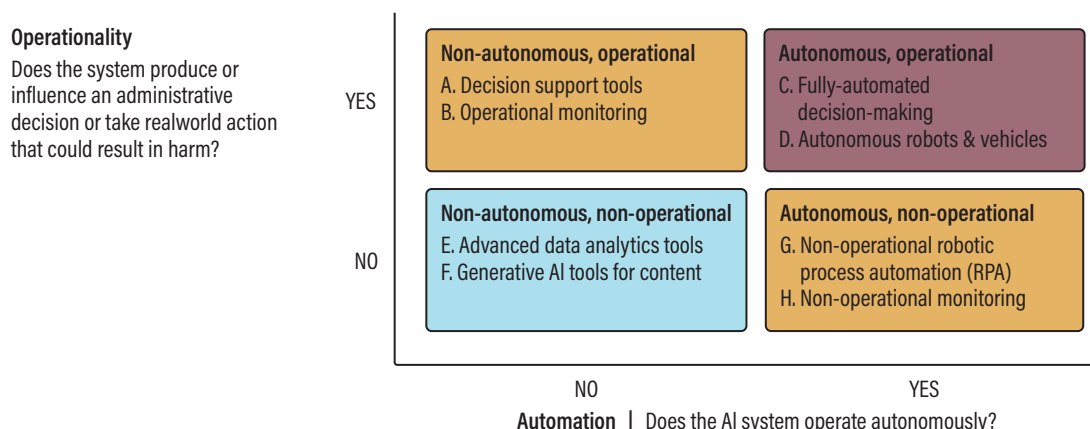
Determining the human rights at risk, and therefore, the potential outcome in the event of poor risk management, in a specific context is made complex by the multiple types of AI, innumerable use cases, and the many underlying mathematical and computational techniques upon which AI systems are based.

Borrowing from the New South Wales AI Assurance Framework terminology and framing, the HTI proposes a categorisation of AI that can serve as a triaging tool in identifying risk. The categorisation distinguishes between two key characteristics of AI systems:

- **Operational:** meaning that it may give rise to a tangible action with a potentially significant effect (e.g. systems for determining home loan or credit eligibility, also facial recognition for retail surveillance).
- **Autonomous:** meaning either that the system operates or is triggered independently (e.g. customer chatbots), or that the system leads to an action or outcome (e.g. self-driving cars), independent of human intervention.

This model can be applied by investors who are interested in determining the scope and scale of human rights violations that may result from the use of particular AI applications as well as assessing the potential outcomes: the greater the degree of operability and automation, the greater the scope and scale of human rights violations. For this reason, the model is one of many useful tools for prioritising among AI systems based on inherent risks requiring mitigation.

FIGURE 1 AI archetypes arranged by operational and autonomous nature



SOURCE:
Human Technology Institute, University of Technology Sydney

CASE STUDY 1: RACIAL BIAS IN HEALTH CARE ALGORITHM



Commercially deployed algorithms are widely used to guide health care decisions globally. These algorithms can cause both opportunities and risks for companies such as insurance companies and healthcare companies.

A study found that a machine learning technology, designed to predict which patients will benefit from extra medical care, dramatically underestimated the health needs of the sickest African-American patients in the United States of America.⁴ Despite the algorithm specifically

excluding race as a data input, it did not account for underlying racial disparities in health costs at a given level of health. The result was an algorithm that was well calibrated by race based on predicted costs rather than actual health needs. Following the study's publication, the algorithm was adjusted to place less importance on expenditure, reducing the algorithmic bias by 84%.

This example highlights the need for AI governance, standards and rigorous testing to support improved patient care and outcomes,

prevent worsening existing health care inequalities, and to avoid potential operational, legal and reputational risks. If not managed appropriately, this may have company valuation impacts. Although the impact of this can be linked to a variety of human rights, two key human rights are highlighted in this example, explicitly, the right to non-discrimination and the right to health of people of colour.

1.3 Emergence of regulatory, governance and ethical frameworks

AI systems are subject to a large body of existing laws. Yet, there are unique issues raised by AI that demand specific regulatory responses. Jurisdictions around the world are taking different approaches to closing these legal gaps: amending existing laws; adopting use-case specific legislation; and adopting broad-based AI legislation, as described below. As the adoption of regulation lags behind advancements in technology, additional approaches include the adoption of ethical frameworks and technical standards to help guide governments and industry as they navigate the ethical and human rights impacts associated with AI. In turn, this also helps investors navigate the potential investment risks.

Australia



Many existing laws apply to the design and deployment of AI systems, including anti-discrimination, consumer protection, cyber security, intellectual property, occupational health and safety, privacy and tort law. Yet, there are gaps in these laws when it comes to the unique risks posed by AI.

To start addressing gaps, in 2019, the Australian Federal Government adopted an **Artificial Intelligence Ethics Framework**⁵ designed to guide businesses and governments in the responsible development, deployment and implementation of AI. The framework includes a set of voluntary AI Ethics Principles (Figure 2), which have general applicability for businesses across the digital technology lifecycle.

In 2024, the Australian Government published its interim response⁶ to the 2023 'Safe and Responsible AI in Australia' consultation. The interim response outlines a less prescriptive approach to AI regulation than that taken in other jurisdictions, instead adopting a "risk-based approach" focusing on the development of specific regulatory safeguards based on high-risk AI use cases and the adoption of voluntary industry standards.

FIGURE 2 Australia's AI Ethics Principles



In considering the right regulatory approach to implementing safety guardrails, the Government's underlying aim is to ensure that the development and deployment of AI systems in Australia in legitimate, but high-risk settings is safe and can be relied upon, while ensuring the use of AI in low-risk settings can continue to flourish largely unimpeded. The immediate focus is on considering what mandatory safeguards are appropriate, informed by developments in other countries.

Aotearoa New Zealand



In Aotearoa New Zealand, the government, together with the World Economic Forum, set out on a project to Reimagine Regulation for the Age of AI⁸. This project sought to co-design actionable governance framework for AI regulation. In line with the AI Strategy for New Zealand, the project thus far has highlighted the need for co-design and flexibility of the system levers, tools and incentives. Specifically, it has “committed to a collaborative partnership with our communities, helping develop their understanding of AI and ensuring Māori values, governance and tikanga are part of our AI ecosystem.”⁹

As highlighted by Mint Asset Management, “the rapid adoption of AI technologies in New Zealand financial organisations occurs in an unregulated environment”¹⁰ as no AI-specific laws exist at the time of writing. AI is currently covered by the 2020 Privacy Act which places personal responsibilities on agencies and organisations for protecting personal information. Further, the Te Mana Mātāpono Matatapu, Office of the New Zealand Privacy Commissioner, offered supplementary guidance on how obligations can be met under the Privacy Act 2020 when using AI tools¹¹.

Europe



The most comprehensive regulatory approach adopted to date, is the **European Union's AI Act**¹² (EU AI Act), 2024, which aims to define clear obligations for both developers and deployers based on the level of risk posed by a given AI system. The EU AI Act has an extraterritorial application in certain circumstances and could foreseeably impact Australian businesses that offer services in Europe and/or to Europeans.

The EU AI Act bans AI tools deemed to carry unacceptable risks, including products for “social scoring” and facial recognition technology in publicly accessible spaces (other than for prescribed law enforcement uses). The Act bans the use of remote biometric identification systems (RBI), that are not used in real-time in publicly accessible spaces (i.e. the random collection and storage of facial recognition data from public internet spaces or CCTV), other than in very narrow circumstances. The Act sets out obligations in the context of other AI systems based on the level of risk they pose, including requiring adequate risk assessment and mitigation systems, and logging of activity to increase accountability and ensure traceability of results.

BEST PRACTICE GUIDANCE



Recognising the limitations of current company disclosures, Australia's national science agency, the Commonwealth Scientific and Industrial Research Organisation (CSIRO), in partnership with asset manager Alphinity Investment Management, has developed a “responsible AI framework” to inform investors on how they can assess their portfolio companies' responsible AI practices. The framework aims to support stronger integration of environmental, social and governance (ESG) and AI-related threats and opportunities into investment practices. It has been developed based on feedback from company engagement and regulatory frameworks such as the EU AI Act.

The framework, company engagement insights and key recommendations will be published in May 2024.

United States of America



At both the State and Federal levels, multiple pieces of legislation and proposals are emerging, including State initiatives focused on algorithmic bias, including specifically in hiring processes, and Federal measures broadly addressing trustworthiness of AI. In October 2023, the **US Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence**¹³ was issued, which is a broad-based effort to guide AI development and deployment through industry regulation, standard-setting and engagement with international partners. The Executive Order directs certain Federal agencies to issue guidance on a range of human rights risks related to the use of AI, including discriminatory recruitment practices and access to housing.

The publication of a voluntary AI risk management framework by the United States National Institute of Standards and Technology (NIST) has been particularly influential. The NIST AI Risk Management Framework 1.0¹⁴ is intended to serve as a resource for organisations that are designing, developing, deploying or using AI systems to help manage the risks of AI. As global and other national standards for responsible AI emerge, the NIST Framework remains influential for its articulation of guardrails for ensuring trustworthiness.

Global Standards



An additional piece in the responsible AI puzzle has been the development of global standards. The International Standards Organisation (ISO) has paid significant attention to the development of AI standards, and most notably ISO/IEC 42001:2023 - AI Management System¹⁵ which, adopted by ISO in 2023 and, in turn, by Standards Australia, specifies requirements for establishing, implementing, maintaining and continually improving AI systems. Organisations can be certified for compliance with ISO/IEC 42001 following successful completion of an audit. This can be a useful reference point for investors when assessing companies and how they apply AI.

SECTION 2: WHY SHOULD INVESTORS CARE ABOUT THE HUMAN RIGHTS IMPACTS OF ARTIFICIAL INTELLIGENCE

AI can increase the quality, efficiency and functionality of products and services and, in doing so, can contribute to the realisation of human rights. However, the magnitude of data required, the opacity of many AI systems, the pressure to increase shareholder returns, and uneven access to technology, can all create risks of harm to people, risks to the reputations of the organisations that build, procure and deploy AI systems, and financial risks to organisations and their investors.

This section outlines some of the key risks and human rights impacts and why they are relevant to investors.

2.1 Reputational and operational risk

The nature and degree of risks posed by AI has the potential to jeopardise the reputation and social license to operate, of companies that fail to mitigate these risks effectively. Reputational and operational concerns are amplified by low levels of public trust in AI. A 2023 global survey conducted by the University of Queensland in conjunction with KPMG Australia, revealed that **26% of respondents did not trust technology companies to develop and use AI in the public interest, and that figure rose to 34% for commercial organisations using AI.**¹⁶

In Australia, where large telecommunication and health insurance companies have experienced major data breaches, the lack of trust reported by respondents was even higher, with the greatest concerns raised about cyber security and privacy.¹⁷

The Australian experience demonstrates the residual costs associated with significant technology failures. For example, the Medibank data breach of 2022 that resulted in sensitive health information of nine million customers being revealed (with significant impacts on affected customers' right to privacy and right to health including mental wellbeing) resulted in executive bonuses being axed. Whilst the attribution to AI is unclear in this example, this led to Medibank's share price dropping 18% following the incident and potential for class action litigation in the future¹⁸. In 2023 a class action lawsuit was filed, adding further impacts to the costs associated with the breach.

The financial and operational costs to businesses of diminished trust and CEO departures can be significant. For example, in 2023, the Chief Executive Officer (CEO) of Optus resigned after a system failure resulted in 10 million people being without access to their phones or the internet, including being able to call emergency services¹⁹. Whilst not directly attributed to AI, the disruption to Optus' services may have impacted individuals' rights to life and health if they were unable to access emergency services.

Another example in the US, in 2024, Congressional hearings were called to "examine and investigate the plague of online child sexual exploitation", according to a statement from the US Senate judiciary committee. The CEOs of five major technology companies (Meta, Snap Inc., TikTok, Discord and X (formerly Twitter)) faced intensive questioning from Congressional members, prompting Meta CEO, Mark Zuckerberg, to issue an apology to parents of children who had died following sexual exploitation or harassment via social media²⁰. This follows a landmark lawsuit, in late 2023, filed against Meta by the attorneys general of 33 states in the US, claiming the owner of Instagram and Facebook purposefully engineered its platforms to be addictive for children and knowingly allowed underage users to hold accounts.²¹ Pending the results of the lawsuit, this could lead to a potential financial impact to all companies involved and which is likely to impact shareholder returns.

2.2 Regulatory risk

Regulatory risk for companies may come from the prospect of policymakers moving to change laws governing AI, in a way that prioritises user safety and upholding human rights. As discussed earlier, the pace of regulatory reform does not keep up with the speed of technology development. When lawmakers finally catch up, there could be negative impacts for companies who may not meet the new and enhanced laws. Significant resources may be required to comply, as was the case with the introduction of European Union privacy laws - General Data Protection Regulation (GDPR) - in 2018²², or the California Consumer Privacy Act of 2018²³ (CCPA). In a similar manner, regulators may introduce guidelines or mandates to ensure fairness, accountability and transparency in AI systems to address ethical implications of automated decision-making. Other areas of regulatory scrutiny are likely to cover intellectual property rights, antitrust and competition and cyber security.

Further regulatory risks to companies deploying AI arise from cross-border regulation. Digital technologies often transcend national borders, posing challenges for companies that need to be compliant in multiple jurisdictions. Conflicting regulations may also hinder international collaboration and innovation. In recognition of the transnational nature of digital technologies, independent online safety regulators from across the world have joined together to create the Global Online Safety Regulators Network (GORSRN). This network brings together independent regulators to cooperate across jurisdictions by sharing information, best practice, expertise and experience, to support coherent and coordinated approaches to online safety issues.²⁴ Australian firms and their directors, also face significant regulatory risk from poorly performing, misused or inappropriate AI systems under existing law. This will likely lead to higher compliance costs for businesses.

Across the world, there has been some litigation challenging the application of AI by reference to human rights law or its local equivalent. As an example, in 2018, Finland's National Non-Discrimination and Equality Tribunal decided that a credit institution's decision not to grant credit to an individual, based on an automated credit rating, was discriminatory and prohibited the credit institution from using this decision-making tool²⁵.

Whilst companies may need significant resources to comply with multiple evolving regulatory environments, non-compliance can result in significant fines and damage to reputation. Thus, it is prudent for investors and companies to be aware of, and seek to comply with, regulatory standards.²⁶

CASE STUDY 2: DATA BREACHES



Data breaches are a common source of financial risk.

In 2013, Yahoo! Inc. experienced multiple data breaches affecting billions of user accounts. The breaches were not disclosed until 2016 and directly impacted the company's acquisition negotiations with Verizon resulting in a price reduction of USD\$350 million (from an original USD\$4.8 billion purchase price) for the acquisition deal²⁹.

In Australia, a breach at Medibank cost the company AUD\$46.4 million in incident response and customer support expenses alone, not including financial penalties as a result of regulatory breaches.³⁰

In 2017 Equifax, a major credit reporting agency, exposed sensitive personal information of nearly 150 million consumers. The company's inadequate security measures and inability to promptly inform the public of the breach, led to a significant drop in its share price (falling by ~15% the day after the breach was revealed³¹). The breach subsequently resulted in a USD\$575 million settlement³².

Notwithstanding the benefits AI can bring to cyber security capabilities, AI systems can still be vulnerable to attacks which, as mentioned in section 1.2, can result in violations of the right to privacy and the release of private personal information, including health and financial information. Cybercriminals can be supported by AI, thereby "reducing the technical know-how required to launch cyberattacks"³³.

2.3 Financial risk

AI failures and/or risks associated with AI can potentially lead to litigation and significant financial losses, through reduced revenue as a result of customer attrition, increased compliance and systems costs, or lower company valuations, if companies are not managing the risks (including human rights) appropriately. For example, in the case of, "falloff in customer or investor trust that could translate into a lower stock price, loss of customers, or slower customer acquisition"²⁷. A further example, in 2023, Google's parent company Alphabet, lost US\$100 billion (or 7-8%) in market value when its new generative AI bot, Bard, produced an inaccurate answer during its first demo²⁸.

Data breaches, which may inevitably be incorporated into AI, can affect the financial value of a company, at least in part due to the risks associated with breaching privacy rights. The costs of remediating human rights impacts associated with cyber security breaches and data leaks can also be significant (see case study 2).

The reputational and regulatory risks already described may result in financial risks, including in the form of:

- **Reduced revenues as a result of customer attrition.**
- **Increased compliance and systems costs to rectify identified weaknesses.**
- **Direct costs of breaches as a direct result of fines, customer remediation and associated legal costs (see examples in the breakout box).**
- **Lower valuations as a result of a loss of confidence amongst investors.**

2.4 Risks of harm to people

In assessing the regulatory and reputational risks, as well as financial risks discussed above, a good starting point is to understand the risk of harm to people, as this will likely underpin regulatory changes in the future.

The risks of harm to people arising from AI exist across the lifecycle of AI-based technologies. This includes conditions of modern slavery in technology supply chains, poor labour standards reported among teams required to curate the data used to design AI systems, social disruption and environmental harms associated with its deployment and ultimately the disposal of technological hardware at scale.

Drawing on HTI's categorisation of risks associated with artificial intelligence³⁴, Table 2 presents examples of the types of harms to people that can be caused during the design and deployment by source of harm. These harms may accrue to particular groups, including the most vulnerable.

TABLE 2 Overview of Artificial intelligence risks: Categories and examples of harm from AI systems (adapted from HTI's categorisation of risks)

Source of Harm	Harm Category	Example
AI system failures	Biased system performance	Gender-biased credit scores
	System fragility	Health system collapse
	Security failure	Personally identifiable data exposed
	AI hallucinations	Harmful content promoted
Malicious or misleading deployment or use	Weaponisation of AI systems	Using generative AI tools to create child sexual abuse material
	Misinformation at scale	Elections and civil and political rights undermined by social manipulation via deepfakes
	Misleading or unfair systems	Recommender system fails to communicate job openings to certain demographics, impacting vulnerable groups' right to decent work
	AI-powered cyber attacks	Personalised phishing emails rob individuals of their livelihoods and impact the right to an adequate standard of living
Overuse, inappropriate or reckless use	Limitations on rights at scale	Erosion of privacy via excessive use of facial recognition
	Economic externalities	Unemployment - with many jobs rendered redundant by technological innovation
	Environmental externalities	Carbon costs and water usage associated with excessive use - contributing to climate change

SECTION 3: INTEGRATION – ASSESSING AI-RELATED HUMAN RIGHTS IMPACTS

3.1 Introduction

The following section aims to provide guidance for investors on identifying and assessing actual and potential adverse AI-related human rights impacts in their investments. The guidance focuses on individual investments, but can be adapted to help engage with asset managers, and can be applied pre- or post-investment.

This toolkit does not aim to offer a prescriptive approach. It is recognised that investors' ESG integration approach/ investment research processes may differ significantly and that different investors may have different views on the financial materiality of these issues.

Nevertheless, a good starting point for investors who are seeking a systemic approach, can be to leverage the existing best practice approach to human rights due diligence based on the UNGPs and existing investor guidance on human rights due diligence, including from UNPRI, IAHR and OECD.³⁵ This framework is described in more detail below for interested investors.

Figure 3 below, provides a high-level outline of an existing best practice approach to human rights due diligence process.³⁶

This section primarily focuses on the first stage of the human rights due diligence process outlined in Figure 3 below: “*identify actual and potential adverse outcomes for people...*”, and covers:

1. Underlying concepts
2. Framework for identifying and assessing AI-related human rights risks/issues
 - a. AI and regulatory context
 - b. Identifying actual and potential adverse human rights risks (i.e. inherent risks)
 - c. Identifying the salient human rights risks (i.e. residual risks)

FIGURE 3 UNPRI human rights due diligence process

POLICY	DUE DILIGENCE PROCESSES				ACCESS TO REMEDY
Adopt a policy commitment to respect internationally recognised human rights	Identify actual and potential negative outcomes for people, arising from investees	Prevent and mitigate the actual and potential negative outcomes identified	Track ongoing management of human rights outcomes	Communicate to clients, beneficiaries and affected stakeholders publicly about outcomes and the actions taken	Enable or provide access to remedy

SOURCE: UNPRI (2023)

3.2 Underlying concepts

The three concepts outlined below fundamentally influence the process for identifying and assessing adverse human rights risks/issues.

- 1. UNGPs Protect, Respect and Remedy Framework:** provides the benchmark of minimum expectations for businesses with respect to respecting human rights.
- 2. Risk to people – a saliency approach:** the UNGPs require businesses and investors, to focus on risk to people rather than the risk to the business.
- 3. Focus on adverse impacts:** this requires human rights due diligence to only consider adverse human rights impacts.

UNGPs Protect, Respect and Remedy Framework³⁷

The process for identifying and assessing actual and potential adverse human rights impacts is grounded in the UNGPs. They represent the authoritative global framework for what is expected from businesses and investors, in preventing and addressing the risk of business causing adverse impacts on human rights from their activities and sets out the responsibilities of businesses in respecting human rights risks related. As part of this there are two key things to note:

1. The UNGP's expect businesses to respect human rights, which is a negative obligation (i.e. businesses are expected to refrain from adversely impacting human rights). While Nation States are expected to protect human rights, which is a positive obligation.
2. The UNGP's expectations for businesses to address adverse human rights impacts is dependent on whether the business caused, contributed to, or is directly linked to, the adverse human impact. See Investor Alliance for Human Rights (IAHR)'s *Investor Toolkit on Human Rights*³⁸ for further detail on this.

Further, given the scope for adverse human rights impacts, the identification and assessment of human rights impacts should be grounded within the international human rights conventions, regulation and standards.

Risk to people - a saliency approach

The approach to identifying and assessing human rights risks, from an investor perspective, is different to how broader investment risks are typically considered:

1. The consideration of the investment risks associated with ESG issues tends to focus on the concept of materiality i.e. whether an E, S or G issue is perceived to have an impact on a company's value.
2. As outlined in the IAHR's *Investor Toolkit on Human Rights*,³⁹ considering human rights impacts, in-line with the UNGPs, focuses on the risk to people first (referred to as the saliency approach). Salient human rights risks/issues are then identified through the severity of the impact and the likelihood of the impact occurring, with the former weighted higher than the latter.

While the saliency approach focuses on the risk to people, material negative impacts on human rights can and do impact company value (refer to Section 2 for examples).

Human rights due diligence focuses on adverse impacts

As noted by the Danish Institute for Human Rights,⁴⁰ the UNGPs explicitly state that human rights due diligence should only consider the adverse human rights impacts from business activities. There are two key reasons for limiting it this way:

1. Including both adverse and positive human rights impacts runs the risk of the human rights due diligence offsetting the negative impacts with positive contributions elsewhere, which the UNGPs make clear is not acceptable.
2. Human rights due diligence is a process designed to enable a demonstration of respect for human rights. This can be achieved by identifying, preventing, mitigating and accounting for how it identifies and addresses, adverse human rights impacts, in this context, the company's accountability based on international expectations, under the UNGPs, to respect human rights.

The consideration of positive human rights impacts should form a separate assessment, which can leverage the guidance provided in Section 3. However, this toolkit does not specifically address identifying and assessing positive human rights impacts.

3.3 Framework for identifying AI-related human rights risks/issues

Below is another suggested approach for investors who, when assessing investment implications, may wish to identify and assess AI-related issues from a human rights risks perspective.

- 1. AI & regulatory context** – understand the nature of the company's activities, supply chain and regulatory obligations as they relate to AI and human rights.
- 2. Identifying actual and potential adverse human rights impacts** – identify potential inherent risks to people driven by key characteristics identified in step 1, leveraging the international human rights framework and guidelines to ensure no adverse human rights impacts are missed.
- 3. Identifying the salient human rights risks/issues** – assess the severity and likelihood of identified adverse human rights risks/issues from step 2 and the existing mitigants to identify the company's salient human rights risks/issues.

This framework is the result of applying the existing best practice approach to human rights due diligence, and the associated underlying concepts, to AI specifically.

AI & regulatory context

The starting point to identifying the AI-related human rights risks for a company is to understand the nature of the company's activities, supply chain and regulatory obligations. In particular understanding:

- **The AI supply chain** - How AI is (or could be) used in the company's activities, including their supply chain and customers.
- **AI System Characteristics** - What are the key characteristics, properties or attributes of the AI system.
- **Regulatory context** - What are AI and human rights-related international human rights standards and regulatory obligations of the company.

Understanding these aspects helps inform the approach to identify and assess the potential adverse human rights impacts.

The AI supply chain

Part of identifying a company's potential AI-related adverse human rights impacts requires understanding of how the company's uses or is connected to the use of AI. Figure 4 provides a simplified overview of the AI supply chain, as outlined by the OECD⁴¹ to assist investors in understanding how a company's activities are related to the use of AI.

Identifying and assessing a company's human rights impacts is influenced by where it sits in the supply chain: developers, vendors or deployers. The following provides an overview of each group, and the relevant issues investors may wish to consider.

AI developers, vendors and deployers

AI developers are involved in the development (and sale) of AI-based products and/or services to their clients, which may have unintended consequences for society. The key considerations focus on whether the developer has responsibly designed, developed, implemented and marketed their product. Issues to consider include:

- Has the developer considered, and appropriately managed, the potential and actual human rights impacts (from design, development and use) of their product?
- What data has the developer used for their product and how was it collected?

AI vendors are involved in the sale of AI-based products and/or services to deployers (end users). The key considerations are whether the vendor has conducted appropriate due diligence on both the developer of the AI products and/or services, and the deployer. Issues to consider include:

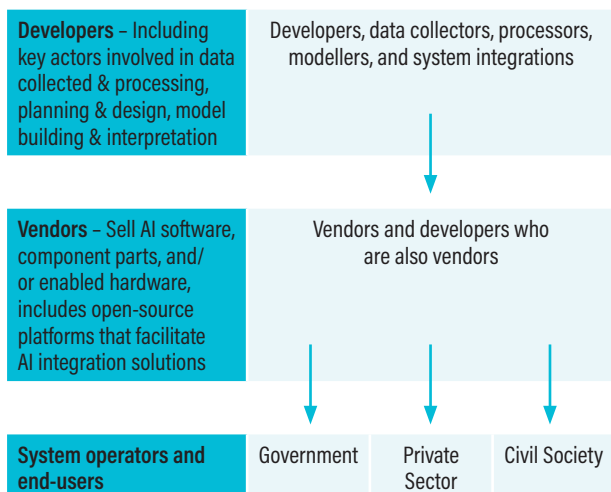
- Does the vendor have a due diligence process for developer (suppliers) and deployers (customers)?
- Does the vendor have the responsibility or resources to assess who they sell to?
- Does the product come with training on AI limitations?
- Where there is significant potential for misuse, does the vendor conduct customer due diligence or implement adequate protections against misuse?

AI deployers (end-users) engage suppliers of a product, service, or solution; often through a tender that includes a due diligence process. The key considerations focus on how the deployer understands and implements the product, service, or solution. Issues to consider include:

- Does the deployer understand what AI is currently in use within its business, and what are the material AI impacts that may contribute to company value and risk?
- Does the deployer have the expertise to understand and identify potential human rights risks/issues?
- Does the deployer have appropriate governance in place to manage the potential human rights risks/issues of the AI product?
- Does the deployer conduct human rights due diligence on the supplier and the AI product itself?

Note: the above apply to deployers of AI who have purchased an AI product or service or created their own AI application.

FIGURE 4 Simplified overview of the AI supply chain from the OCED



SOURCE: OECD

AI RELATED HUMAN RIGHTS DUE DILIGENCE CHECKLIST



See Appendix A for a generic AI-related human rights due diligence checklist.

AI system characteristics

Part of identifying a company's potential AI-related adverse human rights risks/issues requires understanding the underlying characteristics, properties or attributes of the AI system(s) in use. Understanding the key characteristics can be challenging given even the definition of AI is not universally agreed. However, the OECD developed an AI Classification Framework⁴², which provides a useful and consistent framework for understanding the key characteristics of an AI system. The Framework is split into five dimensions:

- People & Planet.
- Economic Context.
- Data & Input.
- AI Model.
- Task & Output.

Each of the Framework's dimensions has a subset of properties and attributes to define and assess policy implications and to guide an innovative and trustworthy approach to AI as outlined in the OECD AI Principles. It is noted that some aspects of the Framework do extend more to identifying adverse human rights impacts.

Further, it is noted that the Framework was developed to help policy makers, regulators, legislators and others characterise AI systems deployed in specific contexts, and to evaluate AI systems from a policy perspective. As AI integrates all sectors at a rapid pace, different AI systems bring different benefits and risks and can be extremely complex. In comparing virtual assistants, self-driving vehicles and video recommendations for children, it is easy to see that the benefits and risks of each are very different. Their specificities require different approaches to policy making and governance, which required a framework to support such policy development.

FIGURE 5 OECD AI Classification Framework

PEOPLE & PLANET	Criteria	Description
USERS	Users of AI system	What is the level of competency of users who interact with the system?
STAKEHOLDERS	Impacted stakeholders	Who is impacted by the system (e.g. consumers, workers, government agencies)?
OPTIONALITY	Optionality and redress	Can users opt out, e.g. switch systems? Can users challenge or correct the output?
HUMAN RIGHTS	Human rights and democratic values	Can the system's outputs impact fundamental human rights (e.g. human dignity, privacy, freedom of expression, non-discrimination, fair trial, remedy, safety)?
WELL-BEING & ENVIRONMENT	Well-being, society and the environment	Can the system's outputs impact areas of life related to well-being (e.g. job quality, the environment, health, social interactions, civic engagement, education)?
<i>DISPLACEMENT</i>	<i>{Displacement potential}</i>	<i>Could the system automate tasks that are or were being executed by humans?</i>
ECONOMIC CONTEXT	Criteria	Description
SECTOR	Industrial sector	Which industrial sector is the system deployed in (e.g. finance, agriculture)?
BUSINESS FUNCTION & MODEL	Business function Business model	What business function(s) is the system employed in (e.g. sales, customer service)? Is the system a for-profit use, non-profit use or public service system?
CRITICALITY	Impacts critical functions/activities	Would a disruption of the system's function/activity affect essential services?
SCALE & MATURITY	Breadth of deployment <i>{Technical maturity}</i>	Is the AI system deployment a pilot, narrow, broad or widespread? <i>How technically mature is the system (Technology Readiness Level -TRL)</i>
DATA & INPUT	Criteria	Description
COLLECTION	Detection and collection Provenance of data and input Dynamic nature	Are the data and input collected by humans, automated sensors or both? Are the data and input from experts; provided, observed, synthetic or derived? Are the data dynamic, static, dynamic updated from time to time or real-time?
RIGHTS & IDENTIFIABILITY	Rights "Identifiability" of personal data	Are the data proprietary, public or personal data (related to identifiable individual)? If personal data, are they anonymised; pseudonymised?
<i>STRUCTURE & FORMAT</i>	<i>{Structure of data and input}</i> <i>{Format of data and metadata}</i>	<i>Are the data structured, semi-structured, complex structured or unstructured?</i> <i>Is the format of the data and metadata standardised or non-standardised?</i>
<i>SCALE</i>	<i>{Scale}</i>	<i>What is the dataset's scale?</i>
<i>QUALITY AND APPROPRIATENESS</i>	<i>{Data quality and appropriateness}</i>	<i>Is the dataset fit for purpose? Is the sample size adequate? Is it representative and complete enough? How noisy are the data?</i>
AI MODEL	Criteria	Description
MODEL CHARACTERISTICS	Model information availability AI model type <i>{Rights associated with model}</i> <i>{Discriminative or generative}</i> <i>{Single or multiple model(s)}</i>	Is any information available about the system's model? Is the model symbolic (human-generated rules), statistical (uses data) or hybrid? <i>Is the model open-source or proprietary, self or third-party managed?</i> <i>Is the model generative, discriminative or both?</i> <i>Is the system composed of one model or several interlinked models?</i>
MODEL-BUILDING	Model-building from machine or human knowledge Model evolution in the field ^{ML} Central or federated learning ^{ML}	Does the system learn based on human-written rules, from data, through supervised learning, through reinforcement learning? Does the model evolve and/or acquire abilities from interacting with data in the field? Is the model trained centrally or in a number of local servers or "edge" devices?
MODEL INFERENCE	<i>{Model development/maintenance}</i> <i>{Deterministic and probabilistic}</i> Transparency and explainability	<i>Is the model universal, customisable or tailored to the AI actor's data?</i> <i>Is the model used in a deterministic or probabilistic manner?</i> If information available to users to allow them to understand model outputs?
TASK & OUTPUT	Criteria	Description
TASKS	Task(s) of the system <i>{Combining tasks and actions into composite systems}</i>	What tasks does the system perform (e.g. recognition, event detection, forecasting)? <i>Does the system combine several tasks and actions (e.g. content generation systems, autonomous systems, control systems)?</i>
ACTION	Action autonomy	How autonomous are the system's actions and what role do humans play?
APPLICATION AREA	Core application area(s)	Does the system belong to a core application area such as human language technologies, computer vision, automation and/or optimisation or robotics?
<i>EVALUATION</i>	<i>{Evaluation methods}</i>	<i>Are standards or methods available for evaluating system output?</i>

NOTE: Criteria and descriptions in grey and marked with an {} symbol = those where objective and consistent information is available. ML = for Artificial Intelligence and Human Rights

SOURCE: OECD

Regulatory context

As outlined in Section 2, regulatory risks vary across jurisdictions and can take the form of legislation specifically focused on the application of AI as well as human rights laws in areas such as privacy, anti-discrimination and critical infrastructure cyber security laws.

While international human rights frameworks do not impose direct obligations, unless they have been enshrined in local legislation, they remain a useful tool in determining which human rights risks are the most salient and support a more people centric approach to due diligence.

3.4 Identifying and assessing human rights impacts

The 5 D's Framework

Identifying the actual and potential human rights associated with AI, which in turn could lead to investment-related risks, can use similar processes as those used to identify the inherent human rights risk in a company. That is, the focus should be on how the AI activities of the company and its value chain may negatively impact individuals or groups.

Figure 6 below is an example of a tool for understanding how AI can impact human rights: “the 5 D's”.

Under the framework, investors could consider the key aspects outlined under each ‘D’ and use the responses to guide the identification of actual and potential human rights impacts.

AI-RELATED HUMAN RIGHTS RISK MATRIX



Appendix B provides a matrix that investors can use to identify the actual and potential impacts of AI human rights risks and how these may result in investment risks. The examples provide colour as to what this may look like in practice, whilst not exhaustive the matrix can help provide some structure to assess the relevant issues and risks.

3.5 Identifying the salient human right risks/issues

A company's salient human rights risks/issues are those that pose the greatest negative impacts to human rights resulting from the company or its value chain's activities. Assessing the salient human rights impacts of a company involves considering three elements:

1. The severity and likelihood of the human rights risks/issues
2. The company's AI & HR governance
3. The company's human rights risk management maturity

Severity and likelihood

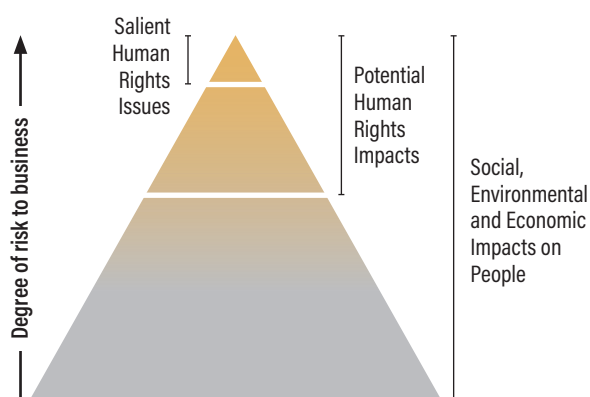
The UNGPs encourage consideration of the severity and likelihood of adverse human rights impacts when identifying a company's salient human rights risks/issues. Figure 7 from Monash University and OHCHR,⁴³ demonstrates the relationship between salient human risks/issues, potential human impacts and broader impacts. Furthermore, it shows that as human rights impacts become more severe, they are expected to be more closely linked to risks posed to the business.

FIGURE 6 The 5D Framework

DATA	DECISION	DISCLOSURE	DISPARITY	DOMAIN
The type of data driving the system, the quality (validity and reliability) of the data as well as the extent to which use of the automated system is consistent with maintaining the integrity (consent, privacy and security) of individuals' personal data.	The nature of the decisions made by the AI system and the extent to which those decisions are made autonomously of human intervention or input .	The extent to which the AI system's inputs, operations and outcomes are transparent, explainable, and contestable by individuals or groups impacted by that system.	The extent to which the AI system may result in an unjust disparity of outcomes between groups, resulting in unfair outcomes or in discrimination or bias against specific individuals or groups based on protected attributes and/or their intersection.	The AI system's scope and localisation of impact concerns matters of legal rights, livelihood, and/or well-being.

SOURCE: KPMG Australia

FIGURE 7 Monash University and OHCHR Degree of risk to business



Assessing the severity and likelihood of the human rights impacts depends on each individual context. However, severity is weighted higher than likelihood.

- Severity is determined by:
 - The scale/gravity of the impact on human rights
 - The scope (i.e. number of individuals) that are or could be impacted
 - The extent to which the impact can be remedied
- Likelihood is determined by:
 - The company's operations
 - The nature of the value chain
 - The presence of vulnerable groups⁴⁴

In addition, an investor may look to the extent to which the company operates in a highly regulated jurisdiction or where companies operate in 'Conflict Affected and High-Risk Areas' or in geographies with a poor human rights track record.

3.6 Assessing risk mitigation

There are a number of factors to be considered in determining the appropriateness of organisational responses and the degree of residual risk. These include governance practices, the efficacy and scope of policies and procedures, capability and resources and internal reporting.

3.7 Good AI governance practice

The HTI report, *The State of AI Governance in Australia*⁴⁵, advocates for a strong governance framework as a positive indicator of a business that considers human rights in its use of digital technology. This can be beneficial for investors who want to protect and enhance overall long-term value for clients and beneficiaries, by engaging with investee companies to encourage best practice.

Outlined below are some indicators of good governance practices:

1. **Director skills matrix:** At a minimum, there should be board level capability in both AI (or digital technology more broadly) and human rights. The Director skills matrix for the Boards of technology solution providers should demonstrate the Directors' relevant knowledge and experience, as well as the organisation's strategy and activities directed at developing and enhancing the Board's understanding of human rights risk management skills. As traditional companies continue to digitalise, technology user organisations could also benefit from having these skills on their Board. This skillset would also be valuable in Chief Executive and other management / leadership roles within a company. Skills should be commensurate to the levels of inherent risk, including considerations of the sensitivity of the data handled, the degree to which AI features within its technology strategy, the potential vulnerability of its stakeholders and the complexity of the regulatory environment in which it operates.
2. **Board committees:** Sub-committees or committees that report up to the Board play a valuable role in directing and informing the Board of planned actions taken to manage human rights risks arising in respect of AI use and development. The Audit & Risk Committee will typically oversee human rights risk management within its remit of overseeing and approving the organisation's existing risk management framework and controls. These frameworks and controls should be adapted to integrate human rights risk, assessed from a "risk to people" lens (rather than the risk to business model which is typically used). This issue may also rest within the Sustainability Committee, especially if the company is at the early stages of understanding their risk exposure and there are activities carried out for such active monitoring and reporting of breaches, 'near misses' and egregious cases. Ultimately, the committee responsible should have formal mechanisms to report to the Board and should delegate day-to-day management of ethics and human rights risks to a role with appropriately seniority and subject-matter expertise e.g. a Sustainability Manager, Ethics Manager, Human Rights Manager or Responsible AI Manager, who reports into the committee and eventually up to the Board and CEO. Furthermore, establishing effective processes for engaging with stakeholders and advisory committees to inform the Board's decision-making on these issues will demonstrate that the Board has access to expert advice and is engaging with stakeholders' concerns.

3. Policies and procedures: Investors may look for policies that establish expectations and behaviour expected of employees and suppliers, as well as processes to manage and deal with human rights issues that emerge. Examples include a Code of Ethics or a Code of Conduct, and specific policies, protocols or frameworks dealing with respecting and mitigating human rights of users and society, based on international human rights standards (e.g. UNGPs), Human Rights, Modern Slavery, artificial intelligence, grievances or remediation or Safety by Design⁴⁶ Risk Assessment. These policies should outline how to manage and deal with human rights issues as they emerge, set principles that guide the organisation's response to human rights issues, and establish the processes that stakeholders can expect will be followed if these issues are raised. Companies with a more mature approach to managing human rights risks may develop a remediation policy and/or plan. Given the continued emergence of issues and the rapid development of technology, issues may not be easily identified in their initial stages, therefore it may be valuable for companies to periodically review the effectiveness of their policies and processes (on an annual basis), identify any new or emerging risk areas, and establish and promote grievance mechanisms to employees, their families, suppliers and other stakeholders, as a way to surface and manager potential human rights risks.

4. Compliance: If a company is not yet impacted by regulatory changes, it may anticipate similar laws and standards to eventually roll out in the markets in which it operates. Companies that seek to align their risk management practices to best practice or voluntary standards will be best placed to minimise any future compliance burden and identify, prevent and mitigate human rights impacts in accordance with law and stakeholder expectations.



RESPONSIBLE AI (RAI) GOVERNANCE INDICATORS

The following table provides an example that can guide conversations with company management and investor relations on AI governance and RAI practices.

Category	Indicator
Board oversight	1 Board accountability
	2 Board capability
RAI commitment	3 Public RAI Policy
	4 Sensitive use cases
	5 RAI target
RAI implementation	6 Dedicated RAI responsibility
	7 Employee awareness
	8 System integration
	9 AI incidents
RAI metrics	10 RAI metrics

Score: X/10		
0-3	4-7	8-10
Weak	Moderate	Strong

For further details please see [Alphinity and CSIRO's report Integrating Responsible AI into ESG: A Framework for Investors](#).

3.8 Assessing the maturity of human rights risk management

Beyond establishing good governance frameworks for the use of AI, best practice is for companies to also operationalise these policies to demonstrate respect for human rights through positive outcomes and/or harm minimisation.

Outlined below are some indicators of practices which may assist investors in assessing the company's human rights risk management maturity.

- 1. Human rights due diligence assessment:** For companies where AI is used for various purposes throughout the whole organisation, conducting a human rights due diligence assessment, and being transparent about the results of this assessment, can be useful to understand how human rights might be impacted through different aspects of the business' operations and value chain, and which of the business' stakeholders may be impacted (e.g. workers in the direct workforce or supply chain, end users or customers).
- 2. Workplace training:** Frequent training of staff and contractors (where relevant), particularly those engaging with AI and digital technology, to identify and assess human rights risks before the impact occurs.
- 3. Grievances mechanisms:** Paired with a whistleblower policy and grievances mechanisms, ideally operated by an independently managed hotline/third-party utilising various channels (email, phone, text), these systems can help companies prevent or mitigate impacts before they occur.

CASE STUDY 3: SPARK NZ (GOVERNANCE AND POLICIES)



Spark NZ, a listed company on the ASX, has made public its governance and policies on AI. Its [Artificial Intelligence Principles](#) are a public commitment to "set out the requirements" for their workforce for when "Spark technologies are designed, deployed and operated" within the business, specifically taking an ethical and responsible approach to the design and operation of AI. This policy makes clear that a human is ultimately accountable in all decision-making, and encourages the use of the Spark whistleblowing process (called Honesty Box) to raise any concerns. The [Spark Human Rights Policy](#) also sets out Spark's commitment to taking a precautionary approach to minimising potential human rights impacts of AI technologies and ensuring AI systems are implemented in a way that is transparent, explainable, subject to human oversight and accountability, and protects human rights.

- 4. Modern slavery and human rights due diligence reporting:** Ultimately, disclosure is key to enabling investors to understand what companies are doing to be responsible in managing the human rights impact of their use and/or development of digital technologies. While investors may wish to see how a company is operationalising its human rights policy or framework, companies may be sensitive about the level of disclosure they provide on their digital technology management. Any disclosures should be made without having to compromise personal information or trade secrets, but under the UNGPs, transparency and reporting is a key requirement, with a proviso that personal data and proprietary trade secrets should be excluded.
- 5. Business insurance:** Investors may be able to identify if a company has recognised the potential risks and impact it may have through insurance coverage and policies.

As there are limitations in disclosure, investors are encouraged to conduct their own research to understand the various applications of digital technology as they evolve and to constantly consider the human rights impacts that they could produce.

3.9 Quantification of potential AI risks (financial and non-financial)

Quantifying the impact of AI risks, particularly in isolation, is a difficult task. Traditional safety metrics allow a company to identify if someone has suffered a mental or physical injury through physical evidence or medical reporting, and record how many people are impacted by safety risks in a specific period of time.

By contrast, AI-related human rights risks are often multi-faceted and harder to quantify. For instance, the contribution of AI driven personalised marketing programs can lead to overspending and financial distress. However, the contribution of a specific campaign can be difficult to determine.

However, there are areas where more specific metrics can be used. For example:

- A high number of reported cyber security breaches, incidents and near-misses can point to a heightened risk of personal data being leaked and available.
- Specific industries that have customers that are minors such as childcare centresⁱⁱ or online websites that enable minors to register, may be able to assess how these breaches, incidents and near-misses heighten risks to children.
- Customer satisfaction surveys assessed in tandem with demographic information and/or industry benchmarks or specific commentary from customers, may indicate discrimination in the use of specific digital technologies e.g. chatbots.

Once the likelihood of a breach is determined, existing examples of the impacts of customer attrition, remediation costs and/or regulatory action could be applied. Similarly these can be used to indicate the risks or impacts on people in the result of incidents occurring.

ⁱⁱ If the childcare centre is uploading photos of children in their care, such personal information gathered online can be misused and result in things like spam, scams, fraud, identity theft or grooming and unwanted contact potentially leading to child sexual abuse online. [Privacy and your child | eSafety Commissioner](#)

Notably, The World Benchmarking Alliance (WBA) has published the *Digital Inclusion Collective Impact Coalition 2023 Progress Report*⁴⁷, with the latest iteration in November 2023 specifically focused on ethics in AI. In addition, *Ranking Digital Rights* assess and ranks large technology and telecommunications companies on their policies and practices in respect of user's rights to privacy and free expression (digital rights).⁴⁸ Both resources can provide investors with a useful guide as to how their portfolio companies perform.

INDIGENOUS DATA SOVEREIGNTY



Indigenous Data Sovereignty relates to the right that First Nations peoples have to manage the collection, ownership and use of data about them, their Country, knowledge and resources⁴⁹. Intrinsicly linked to the right to self-determination as described in the [United Nations Declaration on the Rights of Indigenous Peoples](#), is Indigenous peoples' ownership, control, access, and possession of their data and intellectual property. Advances in AI need to ensure these rights are upheld in the design and deployment. For further resources and guidance related to the application of Indigenous Data Sovereignty please refer to:

- The First Nations principles of ownership, control, access, and possession: [The First Nations Principles of OCAP® - The First Nations Information Governance Centre \(fnigc.ca\)](#)
- The Dhawura Ngilan Business and Investor Guides: [Business Investor Guides | First Nations Heritage Protection Alliance \(culturalheritage.org.au\)](#)
- Indigenous Data Sovereignty: [The legal and cultural considerations \(terrijanke.com.au\)](#)

ADDITIONAL TOOLS



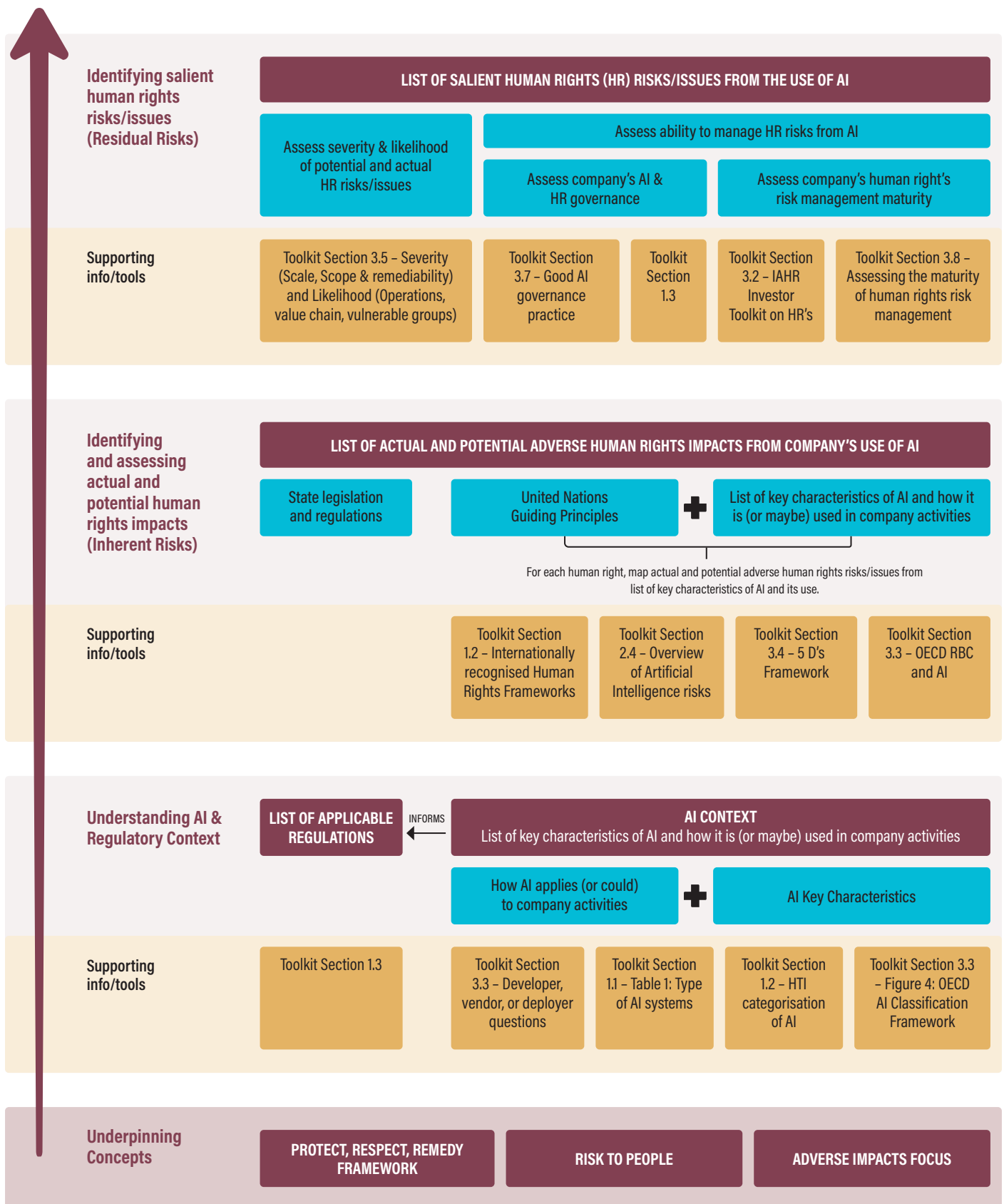
AI in banking: Human Rights Impact Assessments (HRIA)

In 2021 the Australian Human Rights Commission released a report on Human Rights and Technology.⁵⁰ Acknowledging that "new tech should come with robust human rights safeguards", the Commission's recommendations encouraged the need for private sector bodies to undertake HRIA before using AI systems.⁵¹ For this reason, a HIRA tool, in collaboration with NAB was developed by the Australian Human Rights Commission.⁵²

Sustainable digitalisation for the built environment

A multi-disciplinary collaboration drawing on diverse expertise, engaging broadly within Australia and internationally, the [Sustainable Digitalisation Project](#) helps to consider implications of digitalisation and help develop initiatives to put sustainable digitalisation in to practice. The [Sustainable Digitalisation Investment initiative](#) is developing a framework for the application of sustainable digitalisation to real estate and infrastructure asset investment. Aiming to assist real asset owners and managers to: Evaluate the impacts of digital technology-related investment decisions on society and the environment; undertake informed engagements with investee entities to drive positive investment outcomes; and better address associated risks and opportunities, and drive performance down their supply chains.

Section 3 framework graphic



SECTION 4: STEWARDSHIP – HOW INVESTORS CAN ENGAGE COMPANIES ON AI RISKS

Engagement with investee or potential investee companies is an important part of reducing the risks associated with AI. It seeks to communicate the concerns and priorities of investors to a company's leadership, foster better business practices, and hence protect long-term value and returns for clients and beneficiaries.

As outlined in Section 2, companies that fail to effectively manage the risks posed by AI can face significant reputational, legal, and financial impacts.

This section provides guidance for investors on how to prioritise and engage with companies more confidently and effectively on relevant issues, as well as consider other available stewardship tools to encourage better risk management. This is an emerging area, given the rapid pace of innovation and the growth of generative AI. However, investors can set expectations on disclosure and reporting, drawing on the best practice frameworks listed on page 26.

This section covers:

1. Prioritisation
2. Approaches (engagement, voting, divestment)
3. Disclosure & Reporting

4.1 Prioritisation

Investors seeking to prioritise areas for engagement would benefit from first understanding their exposure to a range of adverse human rights impacts from AI as outlined in Section 3, and to consider the resulting risks to supply chain resilience, business stability and reputation with employees and customers. Section 3 provides guidance on how investors may do this.

Once investors understand their exposure to adverse human rights impacts and flow-on risks, investors could prioritise engagement based on their portfolio's most salient human rights issues. This process is the same as the first part of the process for assessing a company's salient human rights (Section 3.3).

EU AI ACT



Investors focused on AI can also refer to the [EU AI Act](#) categorisation as this can have a direct financial impact on the company if mismanaged. EU AI Act (Appendix Y1, 2): Given the potential for fines or banning of business models, investors can prioritise Unacceptable and High-risk use cases in the EU, noting other jurisdictions are taking a different approach.

In addition, some further considerations for prioritising engagement could include:

- Level of leverage and/or ability to influence change within the company
- Whether there are engagement efforts underway by other investors or organisations
- How crucial the company is for the investor (e.g. does the company represent a material portion of the investor's portfolio or how likely is it to in the future)
- Availability of resources for the engagement

4.2 Engagement approaches

Direct or collaborative engagement

Investors can pursue engagement either directly, or through collaboration. Consideration of the best method will be determined by each investor considering their holding and access to the company. Large companies often have extensive shareholder bases, comprised of entities with relatively small percentage holdings. Often, collaborative engagement can result in improved access to these companies and in some cases, more effective engagement outcomes.

A suggested approach in this toolkit would be, after prioritising companies based on the risk framework outlined in Section 4.1, to develop an engagement program – this could be proactive, reactive or both. Proactive engagement plans will be focused on governance, policies, best practices and investment in systems and employees to design, assess and monitor use cases. A reactive engagement strategy may be to engage with a company following an allegation or controversy (e.g. media, court case).

ENGAGEMENT GUIDES



Appendix A provides investors with a generic AI-related human rights due diligence and stewardship guide including example questions for use in engagement. Additionally, Appendix C provides a specific human rights engagement guide that investors can use to develop their engagement plans.

Investors can also look to industry collaborations, for example:

- United Nations Principles for Responsible Investment (PRI) [All Collaborations | PRI \(unpri.org\)](#)
- Investors Alliance for Human Rights: [Digital Rights and AI accountability investor engagement](#)
- World Benchmarking Alliance: [Collective Impact Coalition for Ethical AI](#)
- Other member collaborative partnerships e.g. Federated Hermes Eos and Australian Council of Superannuation Investors (ACSI)

Voting

Voting on resolutions at Annual General Meetings (AGMs) is another tool for investors to influence change, prioritise engagement with companies on these issues, and exercise their voting rights. Over the last four years, there has been an increase in shareholder proposals in the United States technology sector on social issues (Graph 1), some with a direct link to investors' expectations with regards to human rights (Graph 2).

Examples of Shareholder Resolutions on human rights in the US technology industry include Shareholder Proposals regarding:

- Algorithm Disclosures
- Report on Government Requests for Content and Product Removal
- the Human Rights Impacts of Facial Recognition Technology
- Targets and Report on Child Safety Impacts
- Report on AI Misinformation and Disinformation⁵³

The increase in social and human rights related shareholder proposals not only provides investors with opportunities to vote on these important issues, but it can also prioritise these topics with companies in engagement.

Divestment

In the event that an investor considers pursuing divestment, the UNPRI provides the following useful list of considerations for investors in relation to divestment, encouraging investors to:⁵⁴

- assess the extent to which engagement has already taken place, and the results;
- consider the reputational and/or legal risks for investors from engaging with national or local authorities in countries or regions where human rights violations are occurring;
- assess how any divestment decision may affect key stakeholders, such as workers and local impacted communities; and
- understand how any divestment decision may affect other business relationships in different geographies.

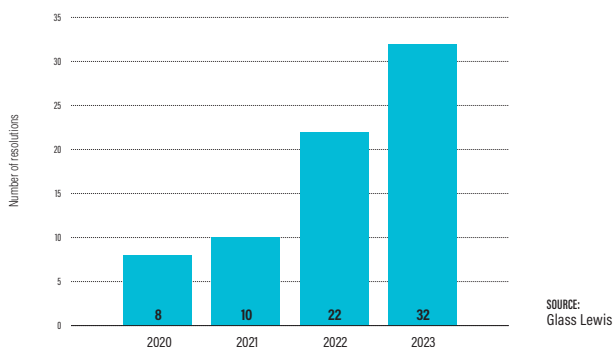
Disclosure and reporting

As the investment-related relevance of these issues will likely accelerate, investors will expect companies to be transparent about how they manage their risks and disclosures and reporting can be an important tool to enable this.

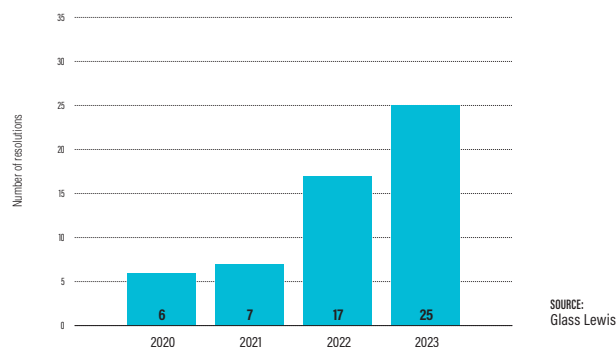
Investors could also consider how they report to their own clients on how they are responding to these issues within portfolios.

'*Transparency and explainability*' is one of the eight pillars of Australia's AI Ethics Principles⁵⁵, a voluntary framework developed by the Commonwealth Government to guide responsible design, development and implementation of AI by business (see figure 2). Responsible disclosure and reporting on human rights risks and impacts associated with AI applications is key to achieving transparency. The AI Ethics Principles make clear that responsible disclosures should be accessible to a range of affected stakeholders and enable them to understand the outcomes of AI systems and the key factors used in decision making.

GRAPH 1 Social shareholder proposals US technology sector



GRAPH 2 Human rights related shareholder proposals US technology sector



RESOURCES TO GUIDE DISCLOSURES



In its November 2023 report 'Advancing Responsible Development and Deployment of Generative AI' the United Nations B-Tech Project emphasised the important role of disclosure and reporting by companies developing generative AI models. B-Tech was established under the auspices of the Office of the High Commissioner for Human Rights (OHCHR) to provide authoritative guidance and resources for technology companies to implement the UNGPs. The Report recognises that robust disclosure and reporting on due diligence, risk assessments and decision-making processes provide critical opportunities for early intervention to prevent or mitigate human rights harms that may arise throughout the AI value chain.

The Report was accompanied by an Overview of Human Rights and Responsible AI Company Practice which noted that while an increasing number of companies publish the Responsible AI (RAI) frameworks, policies and principles that govern their development of AI products and services, to-date few companies have published reports that clearly disclose identified risks to people and society associated with these products and services, and how these risks are managed and mitigated by the company.

In its Generative AI Position Paper, Australia's eSafety Commissioner also encourages companies to prioritise transparency and accountability, noting that service providers should share information with users and regulators about how their models and generative AI systems operate⁵⁶. Under the Online Safety Act 2021, eSafety is empowered to require social media services, relevant electronic services, and designated internet services to report on the steps they are taking to comply with the Government's Basic Online Safety Expectations to make sure these services are transparent, accountable, and safe for people to use.'

Limited corporate disclosure on the impact of AI products and services on human rights hinders investors' ability to identify material human rights risks in their investment portfolio and design metrics to assess the effectiveness of companies' implementation of RAI policies and commitments.

CASE STUDY 4: ENGAGEMENT TO ENHANCE DISCLOSURE: MICROSOFT



One example of good practice disclosure on human rights risks can be found in the Human Rights Impact Assessment of Microsoft's Enterprises Cloud and AI technologies Licensed to U.S. Law Enforcement Agencies. Microsoft engaged a law firm to conduct an independent Human Rights Impacts Assessment (HRIA) in response to a shareholder resolution filed with the company in 2021 which raised concerns that Microsoft could be acting inconsistently with its policies and standards on human rights (including commitments made in its Human Rights Statement) in its contracts and business relationships with Government Agencies. The HRIA considered whether Microsoft's licencing of AI products to law enforcement agencies and immigration authorities caused, contributed or directly linked Microsoft to adverse human rights impacts, particularly for individuals identifying as Black, Indigenous and People of Colour.

Notably, the key findings of that HRIA included that when acting as an "upstream provider of platforms" it was unclear whether Microsoft could be directly linked to adverse human rights impacts downstream.⁵⁷ However, where Microsoft consulted on and participated in the development of AI products for specific uses the HRIA concluded the company would be either directly linked to or contributing to adverse human rights impacts resulting from the downstream use of that product, and therefore may be responsible for mitigating those impacts under the UNGPs.⁵⁸ In accordance with the shareholder request, the HRIA was made publicly available in June 2023.

INTERNATIONAL FRAMEWORKS, STANDARDS AND INSTRUMENTS

Without a global standard framework for companies (or investors) to utilise when considering how to report on AI, here is a list of international frameworks, standards and instruments that may be relevant to reporting on human rights impacts associated with digital technologies that companies and investors could reference and determine suitability. This list is not exhaustive and is designed to provide users of this toolkit with further resources to assist in their activities. Users will need to understand the different frameworks and how they may be relevant to them.

General Disclosures	<ul style="list-style-type: none"> <u>International Financial Reporting Standards (IFRS)</u>, S1 prescribes how companies prepare and report their sustainability-related financial disclosures.
Human Rights Frameworks and Guidelines	<ul style="list-style-type: none"> <u>Universal Declaration of Human Rights (UDHR)</u>: The UDHR, adopted by the United Nations General Assembly in 1948, outlines fundamental human rights principles that are applicable to all areas of life, including the digital realm. <u>International Covenant on Civil and Political Rights (ICCPR)</u>: This treaty, adopted by the United Nations in 1966, addresses civil and political rights, including the right to privacy and freedom of expression. <u>OHCHR Guiding Principles on Business and Human Rights (UNGPs)</u>: The Guiding Principles on Business and Human Rights provide an internationally accepted framework for enhancing standards and practices with regard to business and human rights.
Digital Technology	<ul style="list-style-type: none"> <u>General Data Protection Regulation (GDPR)</u>: The GDPR, enforced by the European Union, is a comprehensive data protection regulation that sets guidelines for the collection and processing of personal data. It enhances individuals' control over their personal information. <u>The Online Safety Act 2021</u>: Provides eSafety with a range of powers and functions to address online safety issues, including those related to generative AI. <u>EU Artificial Intelligence Act</u> <u>National Institute of Standards and Technology (AIRC) Playbook</u> <u>Ranking Digital Rights</u> <u>Interpol Responsible AI Innovation in Action Workbook</u>. This workbook has been developed from a law enforcement agency perspective across a range of human rights principles. <u>Interpol Principles for Responsible Investment – Tool kit</u>

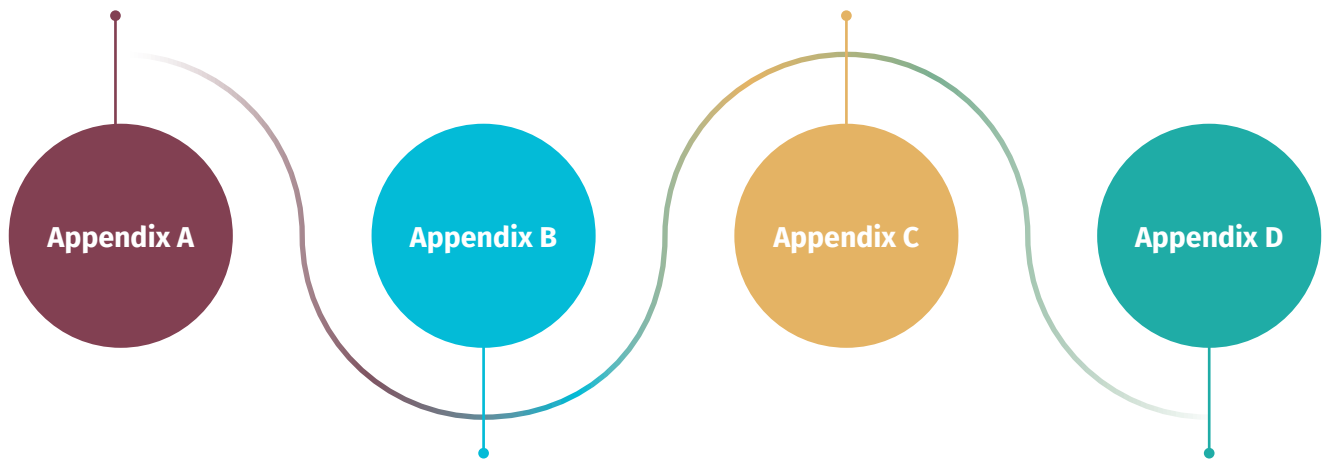
ROAD MAP

Generic AI-related human rights due diligence and stewardship guide

- Your go to risk-based guide aims to provide investors with an initial due diligence and engagement checklist for assessing the AI-related human rights risks and mitigation measures in their assets.

Specific human rights engagement guide

- Additionally, a Specific Human Rights Engagement Guide has been provided that investors could use to develop their engagement plans.



AI-related human rights risk matrix

- This is a matrix that investors can use to identify the actual and potential AI human rights risk impacts and how it may result in investment risks. The examples provide colour as to what this may look like in practice and may draw on issues that have been reported on to the public. It is important to note that the list is not exhaustive but can help provide some structure to assess the relevant issues and risks.

Further resources

- With thanks to Altioem, here you will find links to a list of resources to further your understanding in relation to AI and Human Rights.

APPENDICES

Appendix A: Generic AI-related human rights due diligence and stewardship guide

This “risk-based” guide aims to provide investors with an initial due diligence and engagement checklist for assessing the AI-related human rights risks and mitigation measures in their assets. Investors are encouraged to draw on the other industry specific resources cited in this toolkit to help develop their own due diligence and engagement processes, suited to their specific values, investment philosophy, risk appetite and ESG integration and stewardship objectives.

Asset classification

1. Does the company provide AI-related products and services to any of the following industries:
 - Biometrics, critical infrastructure, education, recruitment, access to essential services (public and private), finance, law enforcement, immigration, administration of justice and democratic processes.
2. Does the company:
 - a. Develop and sell digital technology products that rely significantly on AI?
 - b. Develop its own AI solutions to deliver its core business?
 - c. Use generic AI-related products in the delivery of its core business?
 - d. Use AI products and services in a way that is peripheral to its core business?

An asset/company included in 1 and 2 a, b & c above indicates that a more detailed Human Rights Impact Assessment may be required.

Governance

1. Does the company have an overarching human rights framework and/or policy aligned with the UN Guiding Principles on Business and Human Rights, and which makes reference to respecting AI-related human rights risks and impacts? Or does the company have a policy or framework which explicitly mentions human rights risks?
2. Does the board have oversight over the measures the company has put in place to assess and mitigate the potential AI-related human rights risks and impacts associated with the use of their products and services?
3. Does board membership include a least one person with strong expertise and experience on issues relating to AI and its potential impacts on human rights, and appropriate approaches to mitigation and remedy?
4. Does the company have a board approved digital technology policy (or similar) that includes reference to the assessment and mitigation of AI-related human rights risks and impacts?
5. Is the consideration of technology related, including AI-related human rights risks, included in the charter of a board committee e.g. Audit, Risk & Compliance Committee or similar?
6. Is the assessment of measures to prevent and remedy AI-related human rights risks included in the scope of the internal and external audit mandates?
7. Does the board receive formal reporting on the progress of measures to protect and remedy AI-related human rights risks and incidents?

Management & controls

1. Are the consideration of AI-related human rights risks included in the company's digital technology/AI strategy?
2. Does the company include reference to principles for protecting user rights in its mission statement, corporate values and/or strategy?
3. Does the company have a dedicated advisory or management committee responsible for AI, that includes the responsible management of AI? Does this committee report directly to the board or a board committee?
4. Are AI-related human rights risks specifically included in the company's risk management framework, risk register and risk matrix.
5. Does the company have a human rights due diligence process to assess its real and potential human rights risks and impacts associated with the use of AI in its products and services?
6. Are digital, including AI, solutions designed and deployed in accordance with applicable laws in all relevant regulatory regimes?
7. Does the company have specific digital technology/AI policy guidelines and procedures that provide appropriate guardrails for staff as to the use of AI in product development, sales and operations? For example: eSafety's Safety by Design approach (particularly for the protection of children and vulnerable groups), internal and external testing before the release of an AI systems and ongoing auditing; incorporating digital labelling or 'watermarks'; and mandating 'human-in-the-loop' requirements etc.

8. Does the company's procurement policy and processes include the consideration human rights risks of AI products and services?
9. Is there a specific senior management member appointed and accountable for the company's approach to managing and mitigating AI-related human risks in its operations, products and services and value chain?
10. Are there formal reporting structures, including "whistleblower" procedures in place for staff and teams to escalate concerns to senior management and the board?
11. Does the company provide or participate in an effective grievance and remedy mechanism that are accessible to individuals and communities at risk of harm attributed to the company's AI-related products and services and/or the use of AI in its products and services within its value chain?
12. Is the company's management and/or deployment of AI aligned with and certified to ISO/IEC 42001:2023 – AI Management System?

Products & services

1. Does the company have AI-related "safety by design" principles embedded into its product development processes?
2. Are teams adequately resourced to address AI-related human rights risks associated with the development and sale of their products and services?
3. Has the company complied with all applicable legislation and regulations relevant to product development and sales processes?
4. Who are the intended AI target users and are they likely to be negatively impacted?
5. Are products tested prior to release and are they auditable to identify unintended negative AI-related human rights impacts?
6. How is personal data stored and are the company's privacy and cyber security procedures adequate?

Capabilities

1. Do board members, senior management and staff have appropriate skill sets and experience to enable them to effectively assess and mitigate the ethical and human rights risks of the company's digital technology/AI products and services across the value chain?
2. Are staff and board members adequately trained (i.e. ethics, regulation, human rights, privacy etc) to assess and mitigate AI-related human rights risks and impacts associated with the development and sale of their products and services?
3. Is the meeting of digital technology/AI compliance with human rights norms included in senior management performance KPIs?
4. Are the AI-related human rights risks included in staff/product training resources and are staff aware of their role in preventing and remedying these risks?
5. How does the company foster a culture of accountability and awareness regarding human rights considerations throughout its technological development and deployment processes?
6. Are whistleblower staff trained in AI-related human rights risks and respond to queries/escalate properly?
7. Does the company participate in any industry collaborations pertaining to the prevention and remedying of AI-related human rights risks?

Disclosures

1. Does the company regularly assess and disclose publicly, in alignment with leading practice, information about the effectiveness of its mechanisms to assess, mitigate and remedy the AI-related human rights risks associated with its products and services? Is the company using relevant standards for these disclosures?
2. Are the company's public disclosures in AI-related human rights risks and measures externally verified?

Appendix B: AI-related human rights risk matrix

As per Section 3.4, this is a matrix that investors can use to identify the actual and potential AI human rights risks/issues and how these may result in investment risks. The examples highlight what this may look like in practice and may draw on issues that have been reported on to the public. It is important to note that the list is not exhaustive but can help provide a foundation from which to assess the relevant issues and risks.

Potential rights harmed	Human rights impacts	Investment Risk	Examples
<p>Right to freedom of expression</p> <p>Individuals should be able to express themselves online without fear of retribution or unreasonable censorship.</p>	<p>Automated AI systems are frequently used for content moderation and content curation.</p> <p>These processes may, by design, or unintentionally, limit people's freedom of expression online.</p> <p>Automated AI systems can be used to generate and disseminate large volumes of activity or content that targets a single person.⁵⁹</p>	<p>Reputation risk: negative brand association for companies purposefully or inadvertently censoring content.</p> <p>Regulatory risk: political and government intervention (e.g. government making demands for user information, censorship of content etc.)</p>	<ul style="list-style-type: none"> Using AI bots to carry out volumetric attacks' or 'pile-ons,' that makes it appear there are many people targeting a single person. Algorithms deliberately censoring particular political views or criticisms e.g. Apple not making available certain apps – like VPN access apps on their phones sold in China.⁶⁰
<p>Right to mental and physical health</p> <p>Individuals should not be subject to mental or physical harms as a result of technology design or use.</p>	<p>Lack of responsibility and effective controls with regard to content and data management from companies could affect individuals' mental and physical safety, particularly marginalised groups of people and children.</p>	<p>Reputation risk: negative brand implications if a company supports or facilitates content that result in physical harm, mental health and compromise safety.</p> <p>Regulatory risk: potential for fines, litigation, bans or delays to business models or use of AI (e.g. EU data laws).</p>	<ul style="list-style-type: none"> Social media that allows harmful content to be readily available – especially to minors. Online gambling is extremely accessible anytime/ anywhere. Deepfake AI technology used to generate non-consensual sexual material, with the aim of exploitation, harassment and intimidation, including sexual exploitation of children. Technology can increase the scale and speed with which modern slavery crimes are committed.⁶¹ Automated decision making used by governments to incorrectly target people on income support e.g. Robodebt, Universal Credit. AI-based social scoring. Medical misdiagnosis.
<p>Right to liberty and security of person</p> <p>Individuals must not be subject to arbitrary arrest or detention as result of technology design or misuse of technology.</p>	<p>The personal liberty of individuals can be violated when facial recognition technologies lead to arbitrary arrest and when AI algorithms are used to predict the likelihood that individuals applying for bail or parole will reoffend.</p>	<p>Reputation risk: deterioration of trust of providers and users of poorly calibrated AI technology.</p> <p>Regulatory risk: legal action for unreasonable decision making based on poor algorithms.</p>	<ul style="list-style-type: none"> Authoritarian regimes using data collected by technology companies to target dissidents; regimes shutting down access to internet services to repress opposition. Facial recognition technology used in identification leading to arbitrary arrest and imprisonment.
<p>Right to non-discrimination</p> <p>Individuals should not be subject to discrimination as a result of technology design or misuse of technology and should have equal access to digital technology.</p>	<p>Applications may further exacerbate social inequalities amongst race and gender divides in society.</p> <p>Unequal access to digital technology can create and/or exacerbate divides in society, providing advantages in education, work and society for people who have digital access over those that don't.</p>	<p>Regulatory risk: Legal action, class actions, regulatory fines. E.g. companies not providing equal access may be contravening the Disability Discrimination Act 1992.⁶²</p> <p>Financial risk: delays to product development, increased cost/ investment in employees, training, reporting to meet transparency and disclosure, monitoring including human oversight.</p>	<ul style="list-style-type: none"> Biases within AI used to support credit decisions by banks that unfairly disadvantages specific groups. AI used to support legal decisions. AI used for recruitment or evaluating job candidates, or for monitoring and evaluation of employees that unfairly screens out potential candidates from specific groups. Automated gender "recognition" and AI systems to predict sexual orientation.⁶³ Algorithmic bias arising from incomplete data sets that can discriminate against people based on characteristics such as race, gender or sex. Practices that exploit the vulnerabilities of specific vulnerable groups (e.g. children, persons with disabilities). Discrimination from incomplete datasets (e.g. recruitment, law enforcement or migration).

Rights of the child

Children should be protected from the potential harms of technology design or use.

Potential mental and physical safety risk to children accessing digital technology.

Proliferation of, and access to, content that normalises the sexualisation and abuse of children.

Reputation risk: Negative brand implications.

Regulatory risk: Legal action; developments in regulation e.g. industry codes and standards developed under the Online Safety Act 2021 which outline measures to deal with Class 1 content including child sexual abuse material.⁶³

Financial risk: Redesign of products to restrict children's access; increased cost/investment in employees, training, reporting to meet transparency and disclosure, monitoring including human oversight.

- Social media companies facilitating anonymous bullying, soliciting or grooming of children.
- Content moderation: display of illegal content including crimes and exploitation of people, including children.

Right to privacy

Individuals should retain control of personal information is handled and that it should be securely stored.

Companies' lack of responsibility and effective controls with regard to providing secure access to individual personal data may lead to negative financial implications including theft, reputational damage to an individual and fraud. Resultant impacts to individuals may include loss of assets, deterioration of family relationships and negative mental health consequences.

Reputation risk: Negative brand implications may impact customer attraction and retention and challenge social license to operate.

Regulatory risk: Current regulation and regulatory developments e.g. Privacy Act and related duties.

- General issues around consent for data use/collection/removal - i.e. difficult to delete content on Meta etc.
- Consumer knowledge of what data is collected and how it is used – hidden risks in the event of cyber security event for investors.
- Data used to accelerate or increase harmful behaviour (e.g. Marketing of unhealthy products etc).

Right to remedy

When a digital system unlawfully or disproportionately limits human rights, individuals should have access to an effective remedy.

Effective remedies for human rights breaches fall under the accountability principle, including administrative and judicial mechanisms to address human rights violations.

Reputation risk: social licence and erosion of trust.



Financial risk: cost reparations to victims, remediation to business processes, possible fines and legal costs.

- Litigation challenging the application of AI by reference to human rights law or its local equivalent.⁶⁴

Appendix C: Specific human rights engagement guide

These guidelines are designed to assist investors develop engagement plans more specifically focused on the rights of people, adopting the concept of salience.

Please note that, while focused on AI, these questions can also be utilised for human rights issues relating to the application of digital technology beyond AI as appropriate to the target company.

Potential rights harmed	Example objective/s	Questions
 <p>Right to freedom of expression</p>	<p><i>Governance</i></p> <p>The company policies and processes include protecting freedom of expression of users interacting with the companies' digital technologies.</p>	<ul style="list-style-type: none"> • How do you select what content is censored or removed from your platform? Is there a public charter or risk guide that assists in decision making? • What frameworks and/or legislation has guided this policy position? • Can you provide insights into the company's commitment to fostering an open and inclusive online environment while addressing any challenges related to freedom of expression that may impact its operations and reputation? • How does the company ensure the safety of users to freely express themselves without fear of harassment or intimidation? Can you provide insights into the measures and policies in place to foster a secure online environment that allows for open expression while mitigating risks associated with potential harm or abuse? • Digital Technology users specific question: How does the company prioritise the safety of users to ensure that the platform promotes freedom of expression without exposing individuals to potential harm or harassment? Can you share insights into the specific features, policies, or practices in place to create a secure and supportive online environment for users to express themselves openly?
 <p>Right to mental and physical health</p>	<p><i>Governance</i></p> <p>The company's effectiveness of processes and policies in mitigating mental health/safety harm caused by its AI/digital technology.</p> <p><i>Social</i></p> <p>Assess the extent of a company's dedication to ethical technology development and the well-being of its users.</p>	<ul style="list-style-type: none"> • What are the harms to mental health and risks to safety that could potentially arise because of technology design/misuse? • How does the company ensure that its technology is developed responsibly and ethically, with measures in place to protect users' mental well-being? • How does the company embed protection of users' right not to be subject to mental or physical harm in the design of its technology? • Can you elaborate on the strategies and safeguards in place to prevent adverse effects on users' well-being, demonstrating the company's commitment to responsible and ethical technology development? • Are official safety review procedures part of the product design process? • Do you moderate behaviour on your platform or service (either in-house, outsourced, community-driven or a hybrid approach)? • Can you detect and flag illegal behaviour and content to prevent harm quickly/before it occurs? • If illegal activity occurs, do you have processes in place to notify law enforcement, support services and illegal content hotlines? • Do you have systems in place to identify and act on breaches of your terms of service or community guidelines? • Do you provide users with tools and features that allow them to manage their own safety? • Do you have visible and simple reporting systems and appeals processes for users to lodge complaints or concerns about their safety, which are actioned within dedicated timeframes? • Do you take proactive steps to inform users about safety policies, features and advice on your service? • Where does the company publicly share information relating to user safety?
 <p>Digital inclusion and access</p>	<p><i>Governance</i></p> <p>The company's strategies for promoting inclusivity and eliminating barriers for safe participation online.</p>	<ul style="list-style-type: none"> • How does the company prioritise ensuring unrestricted and equal access to digital infrastructure and technologies, particularly in terms of internet access? Can you provide insights into the company's strategies and initiatives aimed at promoting inclusivity and eliminating barriers to digital participation for individuals? • Acknowledging the potential societal divide caused by unequal access to information via digital technologies, how does the company approach mitigating the human rights risk of creating disparities in education, employment, and broader societal participation through the design or deployment of technology? • What proactive steps are taken to address potential human rights risks associated with information disparities via AI and/or digital technologies? • How does your company plan to proactively address the imperative of a 'just digital transition' in your global operations, particularly in regions or demographics with lower technological proficiency, to ensure a more inclusive and sustainable digital future? • What specific initiatives and strategies are in place to invest in the upskilling of populations vulnerable to human rights risks related to digital technology (e.g. elderly persons, persons with intellectual disabilities or other user populations with limited technological proficiency), safeguarding against potential societal disparities and positive social impacts?

Right to non-discrimination

Governance

The company prevents discrimination in technology design and use to ensure fairness for individuals using AI and the digital technology and to avoid perpetuating existing inequalities.

- How does the company actively ensure that individuals are not subject to discrimination as a result of technology design or the misuse of AI?
- What specific measures are in place to prevent and address potential biases or discriminatory outcomes in the development and deployment of AI and digital technologies?
- In considering the potential impact of AI on discrimination, how does the company work to prevent further exacerbation of societal harms such as racism, sexism, anti-LGBTIQ bias and classism within society?
- How does the company address bias and discrimination in its design to actively contribute to a more inclusive and equitable society? Can you share insights into the strategies and initiatives in place?
- How is the development and delivery of AI informed by the company's broader DEI strategy?

Child rights

Governance

- The company policies and processes protect children from harm caused by AI.
- The company discloses their assessment of the types of harm, how the company protects child safety, and what measures are in place to prevent risks like exposure to inappropriate content or illegal behaviour.
- The company has appropriate safeguards and oversight to mitigate potential harm to users, including exploitation of children.

- What are the types of potential harm to children and children's rights posed by current or prospective AI, technology and digital systems (used or deployed)?
- How does the company prioritise safeguarding children from potential harm resulting from design or misuse? Can you provide details on the measures and safeguards in place to ensure that the company's digital technologies/AI strategies are developed and deployed with a strong focus on protecting the well-being and rights of children?
- How does the company address the potential negative impact on children's development, safety risks, and the risk of inappropriate content that may contribute to illegal behaviour victimising children, such as domestic abuse, sexual abuse, or child pornography?
- How does the company use age assurance measures to identify child users and apply age-appropriate safety and privacy settings?

Social

- The company transparently provides disclosures around privacy risk management – including take down rates, pertaining to child material before it is being viewed.

Right to Privacy

Governance

- The company has clear policies and processes regarding collecting, storing and or selling data.
- The company provides customers/ clients/ user with the ability to manage their own data.

- To what extent does your company collect, store, use and/or sell data and does this consider the need to respect human rights ?
- How does the company guarantee individuals retain control over the way their personal information is handled and the ability to revoke access? What measures are in place to empower users to control their personal information within your digital technologies?
- How does the company ensure secure access to individual personal data, considering the potentially severe consequences such as theft, fraud, negative financial impacts and significant implications for mental health, family relationships and the potential loss of assets?
- What data management practices does the business follow? For example, do you adhere to the GDPR standards?
- How does the business use new legislative requirements? Specifically, how are the findings applied in markets where regulatory changes may become more stringent in the future?

Social

- The company effectively manages and discloses its approach to privacy risk management, including data security.

Appendix D: Further resources



A list of further resources can be found using the following link as part of the Altiorem library:

**RIAA
AI Resources**

**RIAA
DT Resources**

**RIAA
HRWG
Resources**

Endnotes

- 1 United Nations Human Rights Office of the High Commissioner, 2011, *Guiding principles on business and human rights – Implementing the United Nations “Protect, Respect and Remedy” Framework*, <https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf>.
- 2 Human Technology Institute | University of Technology Sydney (uts.edu.au)
- 3 International Labour Organization, 2024, *Conventions and Recommendations*, <<https://www.ilo.org/global/standards/introduction-to-international-labour-standards/conventions-and-recommendations/lang--en/index.htm>>.
- 4 Sendhil Mullainathan, Zaid Obermeyer, Brian Powers, Christine Vogeli, 2019, *Dissecting racial bias in an algorithm used to manage the health of populations* <[Dissecting racial bias in an algorithm used to manage the health of populations](https://www.science.org/doi/10.1126/science.1262162) | Science>.
- 5 Australian Government: Department of Industry, Science and Resources, *Australia’s AI Ethics Principles* <[Australia’s AI Ethics Principles](https://www.dia.gov.au/ai-ethics-principles) | Australia’s Artificial Intelligence Ethics Framework | Department of Industry Science and Resources>.
- 6 Australian Government: Department of Industry, Science and Resources, 2024, *The Australian Government’s interim response to safe and responsible AI consultation*, <[The Australian Government’s interim response to safe and responsible AI consultation](https://www.dia.gov.au/ai-ethics-principles) | Department of Industry Science and Resources>.
- 7 CSIRO and Alphinity Investment Management, *Responsible Investment AI ESG Framework for investors*, <<https://www.csiro.au/en/research/technology-space/ai/responsible-ai/rai-esg-framework-for-investors>>.
- 8 World Economic Forum, 2020, *Reimagining Regulation for the Age of AI: New Zealand Pilot Project White Paper*, <https://www3.weforum.org/docs/WEF_Reimagining_Regulation_Age_AI_2020.pdf>.
- 9 Ministry of Business, Innovation and Employment, 2021, *Artificial Intelligence and Government*, <<https://iaforum.org.nz/2021/03/19/artificial-intelligence-and-government/>>.
- 10 Gina Delgado, 2023, *The AI revolution – what’s next for the financial services industry?*, <<https://www.mintasset.co.nz/news/the-ai-revolution/#:~:text=Currently%2C%20there%20are%20no%20AI-specific%20laws%20in%20New,collectio%2C%20storage%2C%20access%20and%20use%20of%20personal%20information.>>.
- 11 Office of the Privacy Commissioner, 2023, *AI tools and the Privacy Act: Commissioner issues new guidance*, <<https://www.privacy.org.nz/publications/statements-media-releases/ai-tools-and-the-privacy-act-commissioner-issues-new-guidance/>>.
- 12 European Parliament, 2023, *EU AI Act: first regulation on artificial intelligence*, <<https://www.europarl.europa.eu/topics/en/article/20230601ST093804/eu-ai-act-first-regulation-on-artificial-intelligence>>.
- 13 The White House, 2023, *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, <<https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>>.
- 14 National Institute of Standards and Technology U.S Department of Commerce, <[AI Risk Management Framework](https://www.nist.gov/itl/ai-risk-management-framework), <https://www.nist.gov/itl/ai-risk-management-framework>>.
- 15 International Standards Organisation, 2023, *ISO/IEC 42001:2023*, <<https://www.iso.org/standard/81230.html>>.
- 16 Gillespie, N., Lockey, S., Curtis, C., Pool, J., & Akbari, A., 2023, *Trust in Artificial Intelligence: A Global Study*, The University of Queensland and KPMG Australia, <[trust-in-ai-global-insights-2023.pdf](https://www.kpmg.com/au/en/issues-and-insights/articlespublications/trust-in-ai-global-insights-2023.pdf) (kpmg.com)>.
- 17 Ibid.
- 18 Ayesha de Kretser, 2023, *Medibank sheds \$1.8b as every member’s data accessed*, <<https://www.afr.com/companies/financial-services/medibank-breach-to-hit-substantially-more-customers-20221026-p5bsxs>>.
- 19 Tom Williams, 2023, *Optus CEO Kelly Rosmarin resigns ‘in the best interest of Optus’ following nationwide outage*, <<https://www.abc.net.au/news/2023-11-20/optus-ceo-kelly-bayer-rosmarin-resigns-nationwide-outage/103125462>>.
- 20 Kari Paul, 2024, *Zuckerberg tells parents of social media victims at Senate hearing: ‘I’m sorry for everything you’ve been through’*, The Guardian, <[Zuckerberg tells parents of social media victims at Senate hearing: ‘I’m sorry for everything you’ve been through’](https://www.theguardian.com/technology/2024/oct/17/zuckerberg-senate-hearing-social-media-victims) | US Congress | The Guardian>.
- 21 Kari Paul and agencies, 2023, *Meta designed platforms to get children addicted, court documents allege*, The Guardian, <[Meta designed platforms to get children addicted, court documents allege](https://www.theguardian.com/technology/2023/oct/17/meta-designed-platforms-to-get-children-addicted-court-documents-allege) | Meta | The Guardian>.
- 22 European Union, 2016, *Regulation (EU) 2-16/679 of the European Parliament and of the Council of 27th April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection regulation)*, <<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>>.
- 23 California Civil Code, 2018, *California Consumer Privacy Act of 2018*, <https://leginfo.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5>.
- 24 eSafety Commissioner, 2024, *The Global Online Safety Regulators Network*, <<https://www.esafety.gov.au/about-us/who-we-are/international-engagement/the-global-online-safety-regulators-network>>.
- 25 Kate Jones, 2023, *AI governance and human rights*, <<https://www.chathamhouse.org/2023/01/ai-governance-and-human-rights/06-remedies-ai-governance-contribution-human-rights>>.
- 26 Australia: Act No. 135 of 1992, *Disability Discrimination Act 1992*, 5 November 1992, <<https://www.refworld.org/legal/legislation/natlegbod/1992/en/71499>>.
- 27 McKinsey & Company, 2024, *McKinsey on Risk & Resilience: Managing risks to achieve resilience*, <https://pages.mckinsey.com/rs/473-ZYE-522/images/McKinsey_on_Risk_and_Resilience_Issue_16.pdf?version=0&mkt_tok=NDczLVpZRS01MjIAAAGSGdMiAD4701859cChUMrbv2efnPu2hUvNmFpARIWqAW05VayIuZJ34SdH__15GQzvVBBVTq9v1V7DrhVA8e4VTBIUHXV06MD7S3q47sgK13ZAA>.
- 28 ABC News, 2023, *Alphabet shares dive by \$144b after Google AI chatbot Bard makes error in ad*, <[Alphabet shares dive by \\$144b after Google AI chatbot Bard makes error in ad](https://www.abcnews.com/news/2023/10/17/google-ai-chatbot-bard-error-ad-alphabet-shares/) - ABC News>.
- 29 Trautman, L. J., & Ormerod, P. C., 2016, *Corporate Directors and Officers Cybersecurity Standard of Care: The Yahoo Data Breach*. SSRN Electronic Journal, <<https://doi.org/10.2139/ssrn.2883607>>.
- 30 Chirgwin, R., 2023, *Medibank incurred \$75 million in direct tech costs after Cyber Attack*, iNews, <<https://www.itnews.com.au/news/medibank-incurred-75-million-in-direct-tech-costs-after-cyber-attack-600477>>.
- 31 The Economist Group Limited, 2017, *The big data breach suffered by Equifax has alarming implications*, The Economist, <<https://www.economist.com/finance-and-economics/2017/09/16/the-big-data-breach-suffered-by-equifax-has-alarming-implications>>.
- 32 Newman, J., & Ritchie, A., 2023, *Equifax to pay \$575 million as part of settlement with FTC, CFPB, and states related to 2017 data breach*, Federal Trade Commission, <<https://www.ftc.gov/news-events/news/press-releases/2019/07/equifax-pay-575-million-part-settlement-ftc-cfpb-states-related-2017-data-breach>>.
- 33 Giulia Moschetta and Joanna Bouckaert, 2024, *AI and cybersecurity: How to navigate the risks and opportunities*, <<https://www.weforum.org/agenda/2024/02/ai-cybersecurity-how-to-navigate-the-risks-and-opportunities/>>.
- 34 Nicolas Davis and Lauren Solomon, 2023, *The State of AI Governance in Australia*, Human Technology Institute, The University of Sydney, <<https://www.uts.edu.au/sites/default/files/2023-05/HTI%20The%20State%20of%20AI%20Governance%20in%20Australia%20-%2031%20May%202023.pdf>>.

- 35 Investors Alliance for Human Rights, 2020, *Investor Toolkit on Human Rights*, <https://investorsforhumanrights.org/sites/default/files/attachments/2022-03/Full%20Report-%20Investor%20Toolkit%20on%20Human%20Rights%20May%202020_updated.pdf>.
UNPRI, 2023, *How to Identify Human Rights Risks: A Practical Guide in Due Diligence*, <<https://www.unpri.org/human-rights/how-to-identify-human-rights-risks-a-practical-guide-in-due-diligence/11457.article>>
OECD, 2017, *Responsible business conduct for institutional investors: Key considerations for due diligence under the OECD Guidelines for Multinational Enterprises*, <<https://mneguidelines.oecd.org/RBC-for-Institutional-Investors.pdf>>
- 36 UNPRI, 2023, *How to Identify Human Rights Risks: A Practical Guide in Due Diligence*, <<https://www.unpri.org/human-rights/how-to-identify-human-rights-risks-a-practical-guide-in-due-diligence/11457.article>>
- 37 United Nations Human Rights Office of the High Commissioner, 2011, *Guiding principles on business and human rights – Implementing the United Nations “Protect, Respect and Remedy” Framework*, <https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf>.
- 38 Investors Alliance for Human Rights, 2020, *Investor Toolkit on Human Rights*, <https://investorsforhumanrights.org/sites/default/files/attachments/2022-03/Full%20Report-%20Investor%20Toolkit%20on%20Human%20Rights%20May%202020_updated.pdf>.
- 39 Ibid
- 40 The Danish Institute for Human Rights, 2014, *Human rights and impact assessment – Conceptual and Practical Considerations in the Private Sector Context*, <https://www.humanrights.dk/files/media/migrated/matters_of_concern_huri_and_impact_assessment_gotzmann_2014_0.pdf>
- 41 OECD, 2021, *OECD Business and Finance Outlook 2021: AI in Business and Finance*, <https://www.oecd-ilibrary.org/sites/ba682899-en/1/3/3/index.html?itemId=/content/publication/ba682899-en&_csp_=02d27ef0d7308d76a010fd2a9882228f&itemI=oeecd&itemContenttype=book#figure-d1e6666>
- 42 OECD, 2022, *Framework for the Classification of AI Systems*, <<https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1710842116&id=id&accname=guest&checksum=33ECE941BC53155E59DF0FB561B46EFA>>
- 43 Castan Centre for Human Rights Law, 2016, *Human Right Translated 2.0: A Business Reference Guide*, <<https://www.ohchr.org/en/publications/special-issue-publications/human-rights-translated-20-business-reference-guide>>
- 44 Investor Alliance for Human Rights, 2020, *Investor Toolkit on Human Rights*, <https://investorsforhumanrights.org/sites/default/files/attachments/2022-03/Full%20Report-%20Investor%20Toolkit%20on%20Human%20Rights%20May%202020_updated_0.pdf>.
- 45 Human Technology Institute, 2023, *State of AI Governance in Australia*, <<https://www.hti.org.au/insight-summary/state-of-ai-governance-in-australia.pdf>>.
- 46 eSafety Commissioner, 2023, *Assessment Tools*, <<https://www.esafety.gov.au/industry/safety-by-design/assessment-tools>>.
- 47 World Benchmarking Alliance, 2023, *Digital Inclusion Collective Impact Coalition 2023 Progress Report*, <<https://www.worldbenchmarkingalliance.org/impact/digital-inclusion-collective-impact-coalition-progress-report/>>.
- 48 Ranking Digital Rights, 2024, *Ranking Digital Rights advances corporate accountability for human rights in the digital age*, <<https://rankingdigitalrights.org/>>.
- 49 Dr Terri Janke, Clare McKenzie, and Neane Carter, 2023, *Indigenous Data Sovereignty: The legal and cultural considerations*, <<https://www.terrijanke.com.au/post/indigenous-data-sovereignty-the-legal-and-cultural-considerations>>.
- 50 Australian Human Rights Commission, 2021, *Final Report: Human Rights and Technology*, <<https://humanrights.gov.au/our-work/technology-and-human-rights/publications/final-report-human-rights-and-technology>>.
- 51 Ibid
- 52 Australian Human Rights Commission, 2023, *HIRA Tool: AI in Banking*, <<https://humanrights.gov.au/our-work/technology-and-human-rights/publications/hria-tool-ai-banking>>.
- 53 Investor Alliance for Human Rights, 2024, *2024 Tech Proposals*, <<https://investorsforhumanrights.org/news/2024-tech-proposals>>.
- 54 UNPRI, 2023, *Human Rights due diligence for private markets investors: a technical guide*, <<https://www.unpri.org/infrastructure-and-other-real-assets/human-rights-due-diligence-for-private-markets-investors-a-technical-guide/11383.article>>
- 55 Australian Government Department of Industry, Science and Resources, *Australia's AI Ethics Principles*, <<https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles#:~:text=robust%20and%20safe-,Transparency%20and%20explainability,system%20is%20engaging%20with%20them.>>.
- 56 eSafety Commissioner, 2023, *Tech Trends Position Statement Generative AI*, <<https://www.esafety.gov.au/sites/default/files/2023-08/Generative%20AI%20-%20Position%20Statement%20-%20August%202023%20.pdf>>.
- 57 Foley Hoage, 2023, *A Human Rights Impact Assessment of Microsoft's Enterprise Cloud and AI Technologies Licensed to U.S. Law Enforcement Agencies*, <<https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW16RG2>>.
- 58 Ibid
- 59 Sameer Hinduja, 2023, *Generative AI as a Vector for Harassment and Harm*, <<https://cyberbullying.org/generative-ai-as-a-vector-for-harassment-and-harm>>.
- 60 Rafia Shaikh, 2017, *UN Wants Some Answers From Apple Over Its Decision to Comply With China's VPN Demands*, <<https://wccftech.com/un-apple-vpn-removal-china/>>.
- 61 Australian Government, 2020, *National Action Plan to Combat Modern Slavery 2020-25*, <<https://www.homeaffairs.gov.au/criminal-justice/files/nap-combat-modern-slavery-2020-25.pdf>>.
- 62 Australia: Act No. 135 of 1992, *Disability Discrimination Act 1992, 5 November 1992*, <<https://www.refworld.org/legal/legislation/natlegbod/1992/en/71499>>.
- 63 Daniel Leufer, 2023, *Computers are binary, people are not: how AI systems undermine LGBTQ identity*, <<https://www.accessnow.org/how-ai-systems-undermine-lgbtq-identity/>>.
- 64 Kate Jones, 2023, *AI governance and human rights*, <<https://www.chathamhouse.org/2023/01/ai-governance-and-human-rights/06-remedies-ai-governance-contribution-human-rights>>.

