ACHIEVING EFFECTIVE AI GOVERNANCE

From theory to practice

An investor perspective on managing AI risks

RAILPEN

CONTENTS

AUTHORS

Authors:

This report has been prepared by **Chup Priovashini**, **Rory Sullivan** and **Robert Black** (of Chronos Sustainability) and **Caroline Escott**, **Jasmine Porter** and **Sophie Harris** (of Railpen)

Acknowledgements:

We would like to thank **Becks Goodman**, **Paul O'Donnell**, and the **IT Security Team** (of Railpen) for their contributions to the project and their comments on earlier iterations of this report.

RAILPEN

Railpen's purpose is simple – to secure our members' future.

We are responsible for over £34bn in assets on behalf of more than 350,000 members, providing market-leading pension administration and investment services for defined benefit (DB), defined contribution (DC) and hybrid schemes, including the Railways Pension Scheme – one of the UK's largest, longest established, and intricate funds.

We are built on our heritage. Our history of helping everyday, hardworking people save for their retirement stretches back more than half a century. From our beginnings in 1965 as the pensions office for the British Rail Pension Scheme, we've innovated and evolved to serve the needs of the industry and the financial future of its workers.

We are committed to better member outcomes. Our role is delegated to us by our Trustee, whose unwavering passion to put members' interests first is integral to everything we do – from how we administer the schemes to how we select, structure and size investments to protect and grow the long-term value of members' savings.

We are agile and innovative. Guided by integrity, the importance we place on collaboration, and our member-first approach, we work with policymakers and industry leaders to advocate for a progressive and supportive policy environment that helps us to secure our members' future and shape a more sustainable pensions landscape.

We are different from the norm. We put every pound of profit towards paying members' pensions securely, affordably and sustainably. We make this our mission. Our long-term mindset opens up investment opportunities over multiple decades that enrich communities and the environment, set new standards for innovation and contribute to a resilient UK economy.

We are Railpen.

Find out more at www.railpen.com

CHRONOS
INTELLIGENT SUSTAINABILITY

Chronos Sustainability aims to drive positive impact at scale.

We are a specialist sustainability consultancy with the fundamental objective of delivering transformative, systemic change in the social and environmental performance of key industry sectors.

Chronos Sustainability has deep expertise in responsible investment, corporate sustainability and in stakeholder analysis and engagement.

Find out more about Chronos Sustainability: www.chronossustainability.com

FOREWORD BY RAILPEN

The management of long-term risk and opportunity is fundamental to Railpen's investment approach.

As a long-term investor and a universal owner of assets[a], Railpen has a duty to understand and act upon evolving risks and opportunities that could affect our portfolio companies as well as the trajectory and stability of the national and international economy for decades into the future. Artificial intelligence (AI) – rightly heralded as a significant disruptor to how businesses operate – has the potential to drastically change practices at company-level, and even economy-level, over the long-term. We therefore believe its risks and opportunities must be addressed by long-term investors.

AI is already shaping up to be a key topic for most of our portfolio companies. This report focuses on the downside risks associated with AI, rather than the substantial opportunities it may offer. This emphasis reflects the report's primary objective, which is to develop a governance framework that puts into practice risk controls for the responsible use of AI. Nonetheless, it's recognised that strong governance not only mitigates risk but also positions companies to responsibly harness the potential benefits of AI.

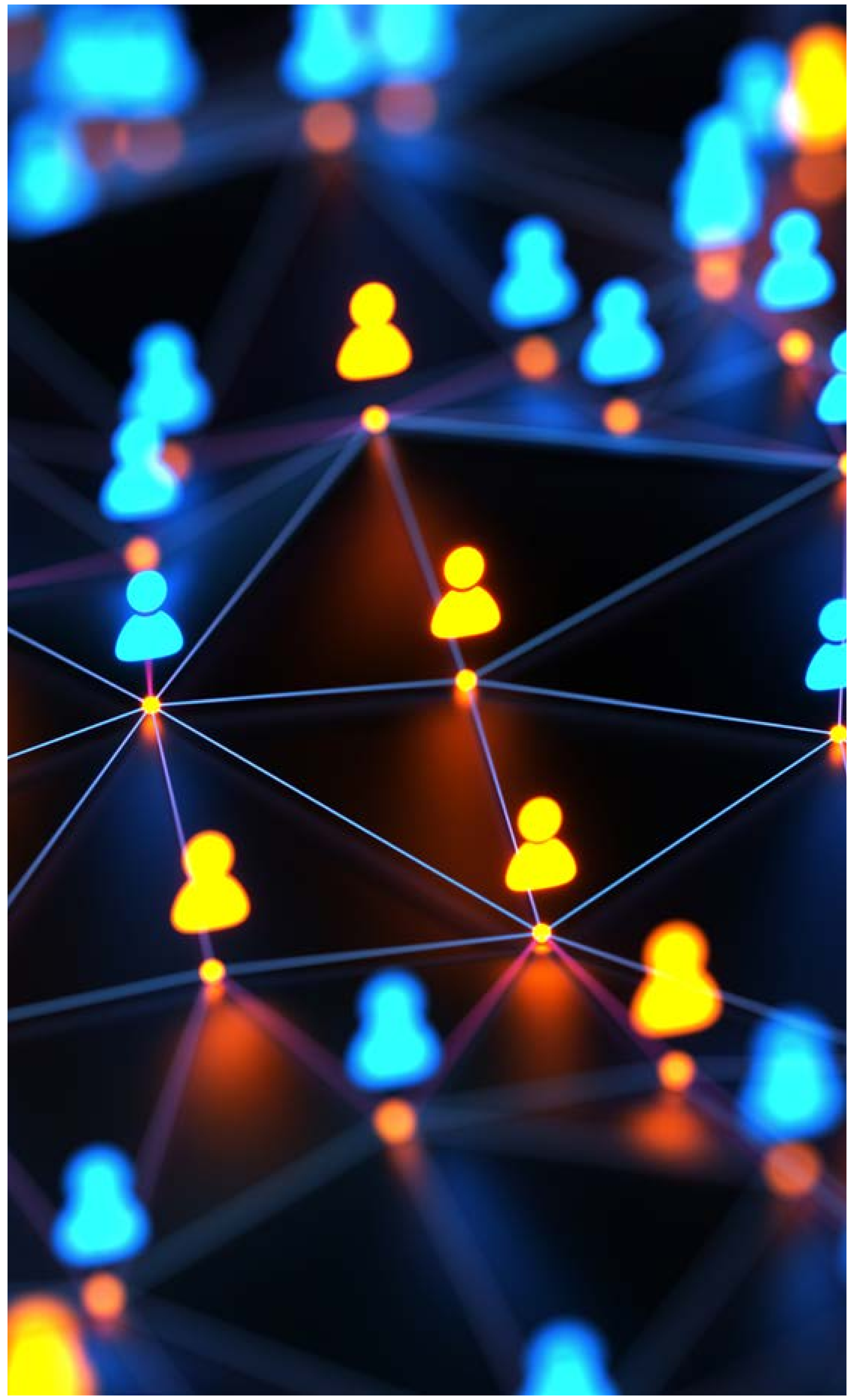
While the speed of AI development means that the long-term capabilities of AI (and long-term risks) are virtually unknown, there's much greater clarity on the short- and medium-term risks associated with AI. These short- and medium-term risks, and how they might be governed and managed, are the focus of this report.

The investor perspective on AI is a useful one. Although investment stewardship practitioners will rarely be AI experts, they bring with them extensive understanding of good governance frameworks and behaviour across a multitude of other issues, as well as (non-commercially sensitive) insights from conversations with other portfolio companies.

Railpen has been strategically evaluating and speaking to portfolio companies on AI over the last few years (see page 24). This report, produced with Chronos Sustainability, consolidates our approach to responsible AI into a coherent stewardship framework and set of activities and provides us with a platform to build on as our knowledge, insight and experience grows alongside our peers and the wider industry.

As well as incorporating many of our recommendations outlined here into our own activities for short- and medium-term AI issues, Railpen also commits to monitoring and engaging with longer-term developments to ensure we are adequately prepared for future progression in this space. Given the system-wide nature of the issues we discuss in this report, we recognise that improvements to our own approach are not enough to achieve the change universal owners of assets need to see.

We have therefore worked with Chronos Sustainability to turn the framework we intend to use ourselves into something that we hope will be of use to investors of all shapes and sizes, as a basis for assessing how companies are managing AI risks.



[a] Pension funds are considered universal owners because they are exposed to systemic risks, such as climate change, inequality, and global pandemics, that cannot be diversified away, and must be addressed through real-world economic change. (Quigley, Universal Ownership in Practice: A Practical Investment Framework for Asset Owners)

EXECUTIVE SUMMARY

AI is becoming increasingly common in business practices, generating significant opportunities as well as material risks across sectors and therefore investment portfolios.

In 2024, 72% of companies had adopted AI in at least one business function, and over 60% of Standard & Poor's (S&P) companies believe they face material risks related to AI[1,2].

This new report 'Achieving effective AI governance' provides an introduction to how AI systems are classified and the risks they pose to portfolio companies. While this report concentrates on the downside risks associated with AI, we recognise the significant opportunities that it presents and believe that strong governance positions companies to both mitigate risks and harness these potential benefits.

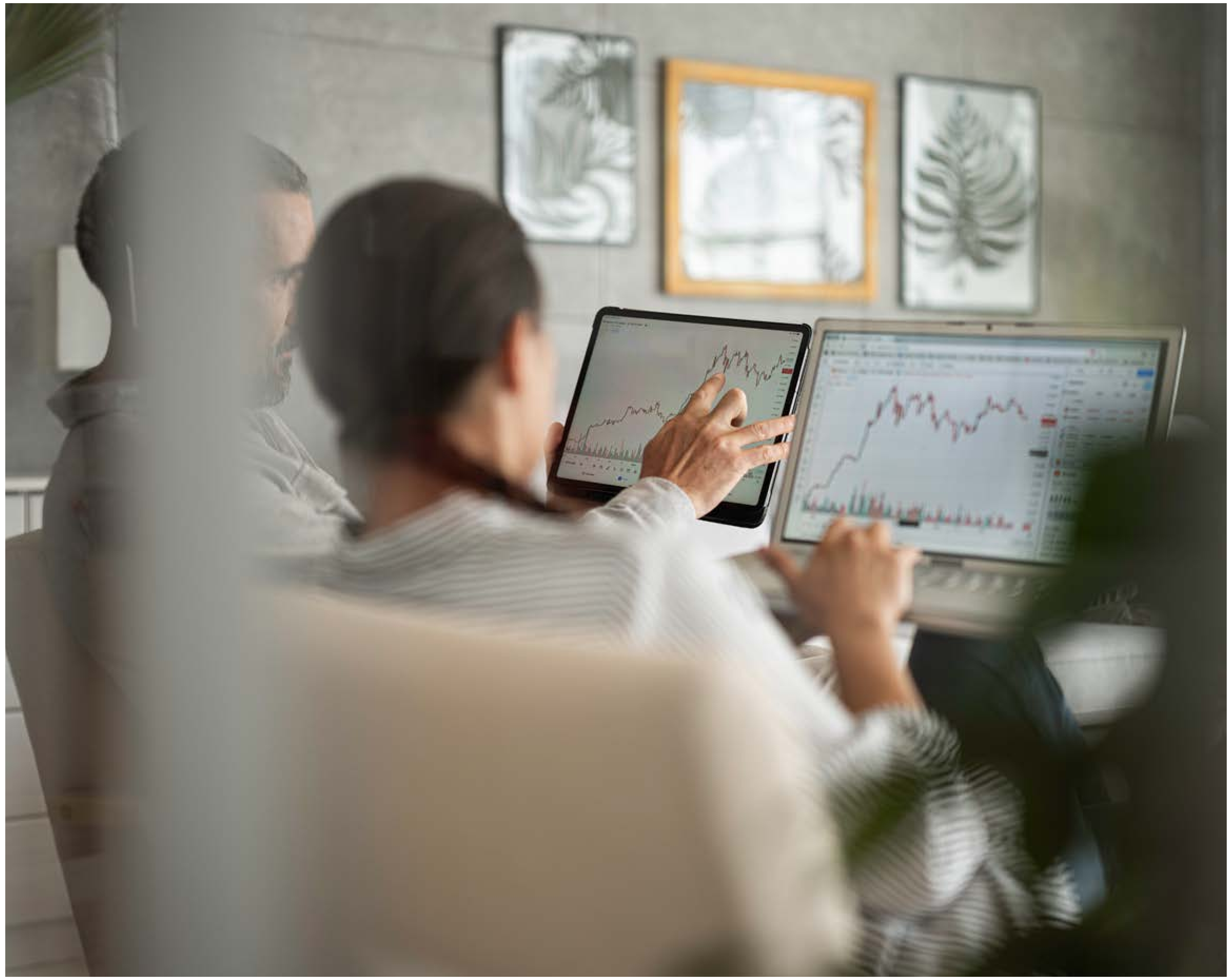
To help investors assess companies' approaches to AI risk management, we also present an AI Governance Framework (AIGF) that translates responsible AI principles into actionable practices. Although the long-term risks, opportunities and capabilities of AI remain largely unknown, the AIGF supports investors' ability to assess portfolio company preparedness for these uncertainties and work more effectively in partnership with them.

In summary, this report seeks to answer the following questions:

- Why should investors be paying attention to AI?
- How can investors assess the significance of AI to their portfolios?
- What can investors do to manage AI risks at portfolio companies?

When answering these questions, we address the activities of both AI developers (companies designing and producing AI) and AI deployers (companies implementing AI) – the roles of which are increasingly blurred, with the latter typically receiving less attention from investors, regulators, and the public despite being exposed to many of the same challenges.

Based on the analysis in this report, we share a series of recommendations for investors that supports collective action in the responsible management of AI.



In particular, we encourage investors to take the following actions:

1. **Conduct a high-level assessment to determine portfolio risks** from AI using the different criteria presented in this report.
2. **Engage with priority companies using our AIGF** as a basis for assessing how companies are managing AI risks, and our guidance on good practice disclosure and potential engagement questions to support meaningful dialogue.
3. **Consider engaging in policy advocacy** around the responsible use of AI to close the gap between regulation and the rapid evolution of AI, and ensure the unique and useful investor perspective is heard by policymakers.

Our AIGF can be categorised across the four pillars (right). As with other risks and opportunities, good governance – that is, the right people supported by effective systems and processes, with timely and transparent reporting to key stakeholders – is fundamental to ensuring companies can effectively consider all related opportunities and risks. With a topic that is developing as rapidly as AI, it is particularly important for an effective governance framework to be in place from the outset.

Our four pillars

Governance

- Oversight of AI
- Management of AI
- AI-related policies

Strategy

- AI relevance to business strategy

Risk management

- Identification of AI risks and impacts
- Management of AI risks and impacts
- Stakeholder engagement

Performance reporting

- Public reporting on AI

INTRODUCTION

AI presents significant opportunities for businesses, with the recent uptake in generative technologies accelerating this trend.

Railpen is actively exploring the long-term opportunities presented by AI, as we see it emerging as one of this decade's most powerful themes (see page 25 and 26). In 2024, 72% of companies had adopted AI in at least one business function[3]. However, as adoption increases, so do the associated risks. A recent survey by Deloitte identified that over 60% of S&P companies believe they face material risks related to AI[4].

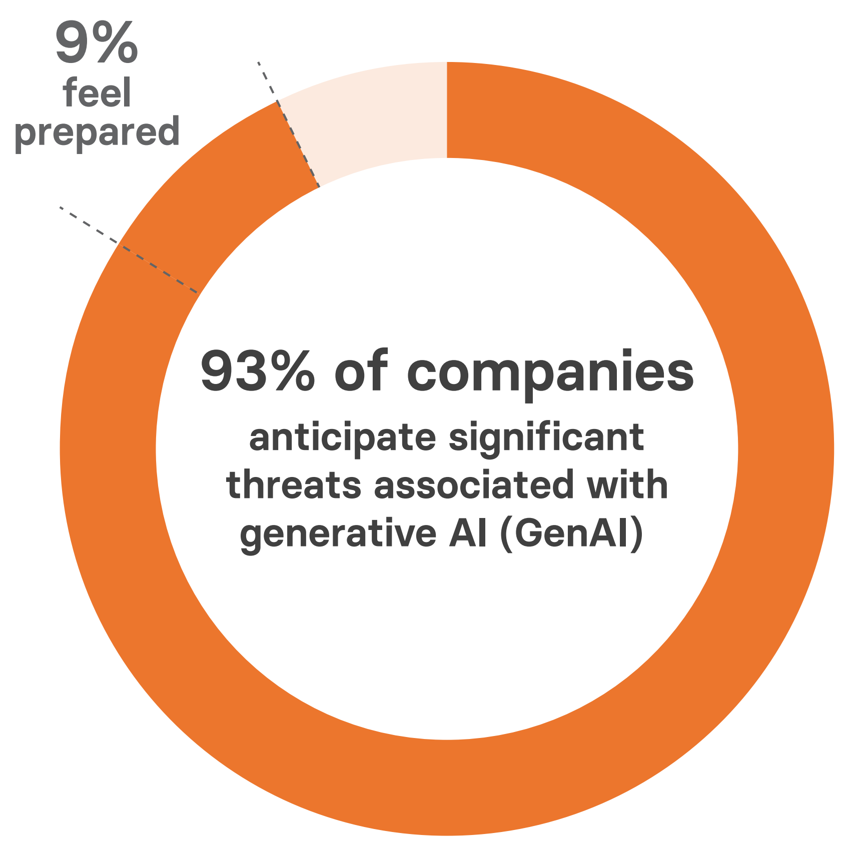
These risks are becoming more apparent as the number of AI-related incidents and controversies steadily increase. In 2023, 123 incidents[b] were reported, marking a 32.3% increase from 2022, with AI incidents having grown twentyfold since 2013[5]. The financial impacts of these risks can be significant, although an extensive and wide-ranging body of empirical research and knowledge around financial materiality is still under development due to the rapid evolution of AI technology. This is demonstrated by the nascent insurance industry in this area and its struggle to price AI risk, given the lack of historical data on AI model performance and the speed at which the technology is changing[6].

Although it is challenging to demonstrate the long-term impacts of AI risk on company valuations, recent incidents illustrate the potential direction of travel. For example, Google's shares fell by 9% in a single day, experiencing a short-term loss of US$100 billion in market cap, after its chatbot Bard made a factual error[7]. In addition, OpenAI was fined €15 million for processing users' personal data without adequate legal justification and by violating the General Data Protection Regulation's (GDPR) principle of transparency[8].

There appears to be a gap between the increasing recognition of AI risks and the governance frameworks established to address them. According to a survey of 300 risk and compliance professionals, **93% of companies anticipate significant threats associated with generative AI (GenAI), but only 9% feel prepared to manage them**[9]. Regulation is not developing quickly enough to address this disconnect. The AI Action Summit held in Paris in February 2025 highlighted a shift in AI policy towards action and opportunities, prioritising deregulation and innovation over safety. We are supportive of policy that helps companies harness these tremendous opportunities, but a balanced approach to risk management should also be encouraged.

Investors are owners of capital, have governance expertise and bring insight and perspectives from conversations with a wide range of AI deployers and developers. Investors therefore play a crucial role in ensuring responsible AI use by companies and can help fill the regulatory gap, at least in part, and encourage effective AI governance.

The AIGF provided in this report is a tool to help companies and investors address current AI risks while also preparing for both the long-term uncertainties and opportunities. Developing robust governance practices can help organisations adapt and be resilient in the face of changes in technology, regulation, and societal impacts[10]. This framework aims to translate responsible AI principles into actionable practices for deployers and developers.



[b] The AI Incident Database defines an AI incident as 'an alleged harm or near harm event to people, property, or the environment where an AI system is implicated'.

WHY SHOULD INVESTORS BE PAYING ATTENTION TO AI?

It is important to define AI as far as is possible, before considering why and how this is a material issue for investors and therefore one that should be prioritised in their dialogue with companies and policymakers.

In this section we set out our understanding of AI, highlight key risks associated with it, both in theory and in practice, and emphasise the need for investors to start deepening their engagement with companies despite having incomplete information and facing significant uncertainty about the long-term direction of travel.

Defining AI

Although there's no standard definition of AI, the term generally refers to a range of technologies designed to mimic human thought patterns in order to solve complex tasks[11]. AI is commonly defined by its processes, as all AI technologies are machine-based systems which use algorithms to interpret inputs (i.e. data) in order to generate outputs (e.g. a prediction, image, or recommendation)[12]. The level of automation within AI technologies varies once they have been deployed.

The aim of AI research is to progressively improve technologies to achieve outputs that are as close to human decision-making and processing as possible. In this context, AI is typically classified by its level of sophistication, looking at its capability or functionality. Current AI technologies are mostly primitive according to these classifications (see appendix tables 1 and 2)[13].

From an investment perspective, a more useful classification is one that focuses on the specific branch or type of technology being applied. As AI develops, research areas have been categorised into branches, differentiated by their goals, applications, functions, and the technologies used to create them. These branches are not standardised or consistently classified, with sources varying considerably on what constitutes a distinct branch, or which determining characteristics (e.g. goals, underlying technologies) should be used to distinguish one branch of AI from another. With this in mind, we describe the commonly identified branches[c] discussed in the context of AI (see appendix table 1)[14,15,16,17].

[c] Other common AI subfields not in the table include:

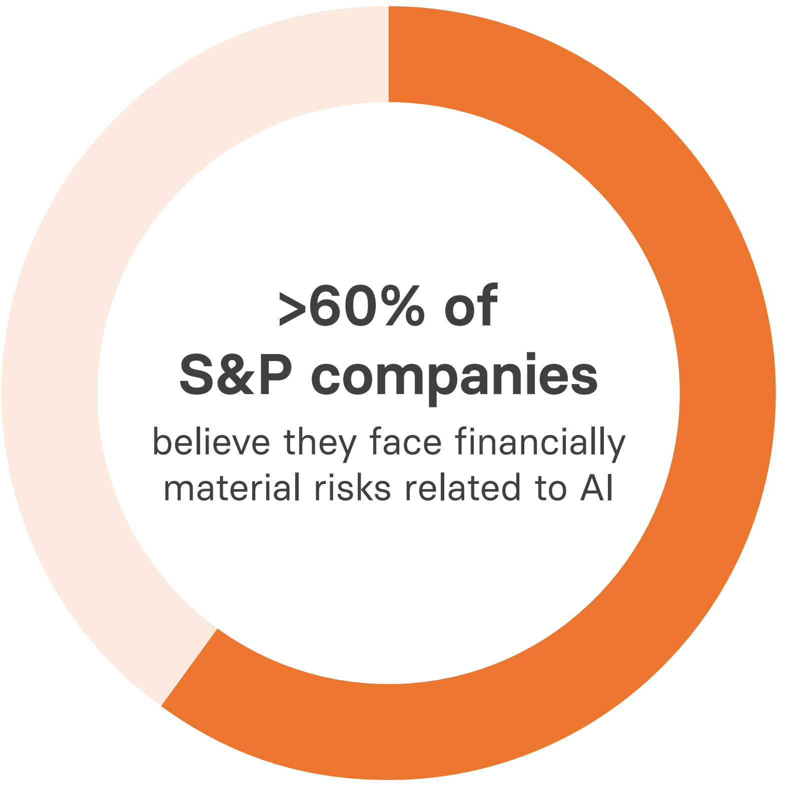
- Fuzzy Logic: systems designed to act based on uncertain or imprecise information employ fuzzy logic, e.g. temperature controls, automated braking systems.
- Automated planning: systems designed to achieve an end goal under given constraints and contexts, e.g. navigation.

Figure 1: AI classification by branch

Branch	Description
Expert systems	Programs specialising in a particular task within one field, designed to recognise patterns or solve intricate problems. Expert systems use rules predefined by humans against a specialised knowledge base.
Machine learning and deep learning	**Machine learning (ML)** algorithms interpret large datasets of known properties (inputs) to generalise, select or predict unknown properties (outputs). ML systems 'learn' over time, i.e. improve on a specific task without explicit instructions. **Deep learning (DL)** is a subset of ML which uses more complex, multi-layered artificial neural networks inspired by the human brain. DL systems require less human intervention because they can label unstructured data on their own and 'learn from their own errors'. Today, ML, including DL, underpins most other branches and is what we generally think of as AI. DL underpins all GenAI.
Generative AI	**Generative AI (GenAI)** uses DL to generate new content which mimics the data it was trained on. GenAI learns a dataset well enough to create entirely new outputs (images, text, audio, and video) that still adhere to the underlying patterns. **Large language models (LLMs)** such as ChatGPT are a type of GenAI built to mimic human knowledge and conversation based on probabilistic models. With the advent of GenAI, all AI may be understood through two distinct categories: **generative** and **traditional**. GenAI produces new and unique content, whereas traditional AI merely interprets the input data to produce predictions, recommendations and decisions.
Natural language processing	**Natural language processing (NLP)** allows computer systems to recognise, process and interpret human language. NLP is used in most text or speech-based AI applications. NLP may be differentiated by its complexity. Simple NLP parses components of language, e.g. syntax, for applications such as spellchecks and translation. Sentiment analysis is an advanced form of NLP used to interpret affective states and subjective information such as humour and idioms.
Perception and computer vision	**Perception** is the ability of systems to perceive visual and audio inputs (e.g. from existing data, a camera, or microphone) in order to analyse, categorise, and derive meaning from them. When applied to visual data (images and videos), perception is also known as **computer vision**.
Robotics	**Robotics** covers systems utilising AI to design, build and program machinery capable of performing tasks autonomously or semi-autonomously. In robotics, sensors gather additional input data from the environment which are used to determine the actions (outputs) performed. Robotics usually employ other AI skills such as NLP or perception to enhance accuracy.

Key AI risks for companies

With over 60% of S&P companies believing they face financially material risks related to AI[18], investors need to understand the different types of risks that companies may encounter to assess whether appropriate governance structures are in place to manage them.



This section outlines the key risks that companies face with the development and deployment of AI. It includes specific examples of AI risks materialising and their financial impact.

Although new models may be trained to reduce the likelihood of known risks identified in previous models, most of the risks are as a result of characteristics that the majority of AI systems share[19,20]:

- **Large data (and energy) requirements:** AI systems must be trained on large volumes of input data (thousands or millions of data points) to form a coherent reference for solving queries. These data requirements make systems energy intensive and vulnerable to issues related to both data provenance and security.
- **Uncritical interpretation of inputs:** All AI systems are conceived, trained and used by humans. Without additional safeguards, they can produce outputs deemed harmful or that reproduce biases present in the training data.
- **Lack of verifiability:** Similarly, AI systems possess no inherent ability to discern between true or false information. Their ability to derive conclusions and retrieve 'facts' relies on inputs, probabilities and associations.
- **Lack of explainability:** AI models are increasingly 'black boxes'. The larger and more complex a DL system is, the more difficult it is to trace the origin of a particular output. Tracing the origin of a particular output is essential for accountability, bias detection, and ensuring that decisions can be audited and trusted.

What is a systemic risk and why does system-wide stewardship matter?

Systemic risks are large-scale threats that cannot be diversified away by individual investors or asset owners, as they affect the entire financial system and economy, and therefore all portfolio constituents.

To address these portfolio-wide risks, investors should consider deploying system-wide stewardship strategies to help understand and mitigate a range of challenges such as climate change, biodiversity loss, wealth inequality – and the rapid development of AI.

The AI value chain involves several types of risks, including those related to risk control, inputs, outputs, and risks to the wider system (known as systemic-risks)[d,e]. Developers and deployers of AI tools are exposed to violations of data privacy laws, litigation due to adverse generated content, and liabilities arising from poorly overseen AI-driven decisions. Other systemic risks, including the depletion of natural resources and occurrence of cyberattacks, can also be amplified by AI[f].

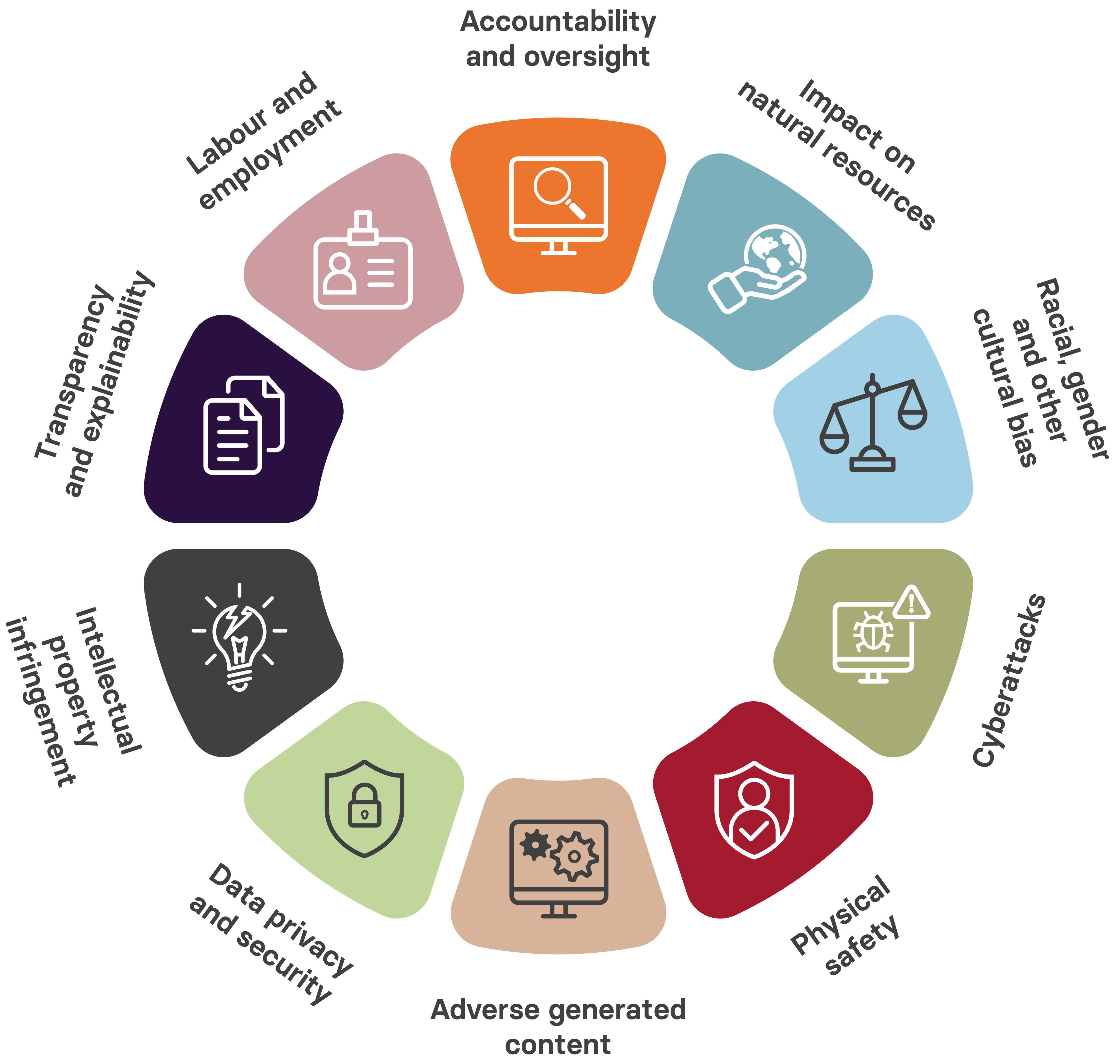
Figure 2 (see page 11) provides an overview of the risks. The information that follows on pages 12-19 provides a more detailed description of each risk as well as the associated potential financial implications.

[d] AI risks can occur at the input, inference, or output stages, but they typically manifest as incidents when the output is generated, meaning most AI risks appear at the deployment stage.

[e] AI incidents usually involve the manifestation of multiple risks at once.

[f] For more insights on how AI is significantly amplifying cybersecurity threats to portfolio companies, please refer to the joint report on Cybersecurity Risk and Resilience by Railpen and Royal London Asset Management.

Figure 2: **Overview of AI-specific risks to companies**


Accountability and oversight
Impact on natural resources
Racial, gender and other cultural bias
Cyberattacks
Physical safety
Adverse generated content
Data privacy and security
Intellectual property infringement
Transparency and explainability
Labour and employment

AI-specific risks and how they can materialise

The following examples (pages 12-19) show how the theoretical risks we've outlined in **figure 2** (left) can play out in practice. We provide an insight into the material impact this has on a wide range of deployers and developers across various industries to the ultimate detriment of shareholders and savers.

These examples are important because they demonstrate why investors need to deepen their dialogue with portfolio companies on approaches to AI, while also beginning to recognise this is a system-wide issue.

Accountability and oversight

Across sectors, AI systems are used to analyse input data to either make or contribute to decisions in areas such as recruitment, customer support, and supply chain management. Physical operations, e.g. self-driving cars, product packaging and handling, or robotic surgery also use AI to make and execute decisions.

Misuse of AI occurs when it is relied on to make subjective assessments, and when AI-assisted decisions are not checked and verified through human oversight.

There are numerous cases where the responsibility for the outcome of an AI decision was unclear or not accepted by the company, leading to debate, distrust and litigation.

Relevant AI branches: All

Relevant business sectors: All

Potential financial impact: Companies who don't maintain effective oversight may face fines and/or criminal charges depending on the severity of the consequences of an AI-assisted decision.

Example incidents:

- CONSUMER DISCRETIONARY
  In 2019, a pedestrian was killed by a self-driving Uber vehicle. The company, which developed and deployed the vehicle, was eventually cleared of criminal wrongdoing, and the vehicle's back-up driver faced charges, a decision that continues to be debated by legal experts[21].
- INDUSTRIALS
  After its AI customer service chatbot gave a customer false information about its services, Air Canada argued the chatbot was "responsible for its own actions". The claim was rejected by tribunal which concluded that companies are liable for what their tech says and does[22].

Impact on natural resources

The speed at which ML-based systems process vast amounts of data means that developing and operating AI systems, especially GenAI, requires immense computational power. This means data centres consume significant amounts of energy and water. While there is a lack of transparency, it is estimated that training a LLM like GPT-3 over millions of Graphics Processing Unit (GPU) hours used the same amount of power consumed annually by 130 US homes[23].

Emissions continue in the use phase, for example, an AI-powered text query uses approximately 10 times more energy than a manual Google search. As GenAI use accelerates, electricity demand from data centres is predicted to increase by 160% by the end of the decade. For example, in the US, they may be responsible for up to 8% of electricity consumption by 2030, compared with 3% in 2022[24]. By 2027, AI's water usage is predicted to hit 6.6 billion m³ [25].

As the need for more data centres increases, the lifecycle of key minerals necessary for hardware (mining and disposal) will create further environmental impacts.

Relevant AI branches: ML, GenAI

Relevant business sectors: IT

Potential financial impact: Companies may face intensifying competition for resources, reduced supply and increased costs of consumption.

Example incidents:

- IT
  Investment in AI has led Microsoft and Google to reconsider their net zero goals. In 2024, Google's carbon footprint increased by 48% since 2019, and Microsoft's increased by 21% since 2021, due to energy use associated with the companies' GenAI development[26]. Both companies previously set commitments to reach net zero by 2030 and are now facing challenges[27].
- IT
  In Chile and Uruguay, the building of new data centres by Google have led to protests against excessive use of freshwater. Communities in the United States and Uruguay have taken legal action to access information on Google's use of local freshwater[28,29].

Racial, gender and other cultural bias

Race, gender, age and other cultural bias in AI models occur because models are trained on large datasets which, by nature, reflect societal biases.

Bias occurs not only at the data input and modelling stage, but also in how projects are designed. ML can reproduce biased associations as well as reinforce and strengthen them – this is because outputs are reintegrated as inputs, causing a feedback loop[30].

Biases may exist anywhere where AI is used to evaluate individuals or groups (e.g. recruitment, policing and criminal justice, and eligibility for insurance and mortgage products). Visual biases occur in image-based AI systems, such as medical diagnosis and photo/art generation.

Relevant AI branches: ML, GenAI, perception/ computer vision

Relevant business sectors: All

Potential financial impact: Companies found to be perpetuating bias may face litigation.

Example incidents:

- IT
  In 2018, Amazon scrapped a recruitment tool that 'taught itself' to disadvantage women. The system was trained mainly on male applicants' data, reflecting their over-representation in the tech sector. For example, it penalised CVs containing the word 'women's' and downgraded CVs from all-women colleges[31,32].
- HEALTHCARE
  AI medical diagnostic systems can have racial and gender biases. For example, skin cancer is under-detected in black patients as training materials use images of mostly white skin[33].
- FINANCIAL SERVICES
  US mortgage lenders using AI underwriting schemes were found to reject applicants from ethnic minority backgrounds 40-80% more frequently than white counterparts with a similar financial background[34].

Cyberattacks

Cyberattacks are an issue across all companies using IT systems but the increasing use of AI significantly amplifies the risks they pose for several reasons:

1. AI requires large amounts of data and therefore more data is at stake in breaches including sensitive data.
2. The use of AI systems opens up the possibility of those systems being exploited by attackers, e.g. through model tampering or polluting the training data, or simply by exploiting blind spots in the algorithm.
3. AI use by malicious actors is increasing, leading to new, sophisticated methods, e.g. deepfakes and faster, less detectable executions of existing methods such as password cracking and data exfiltration[35].

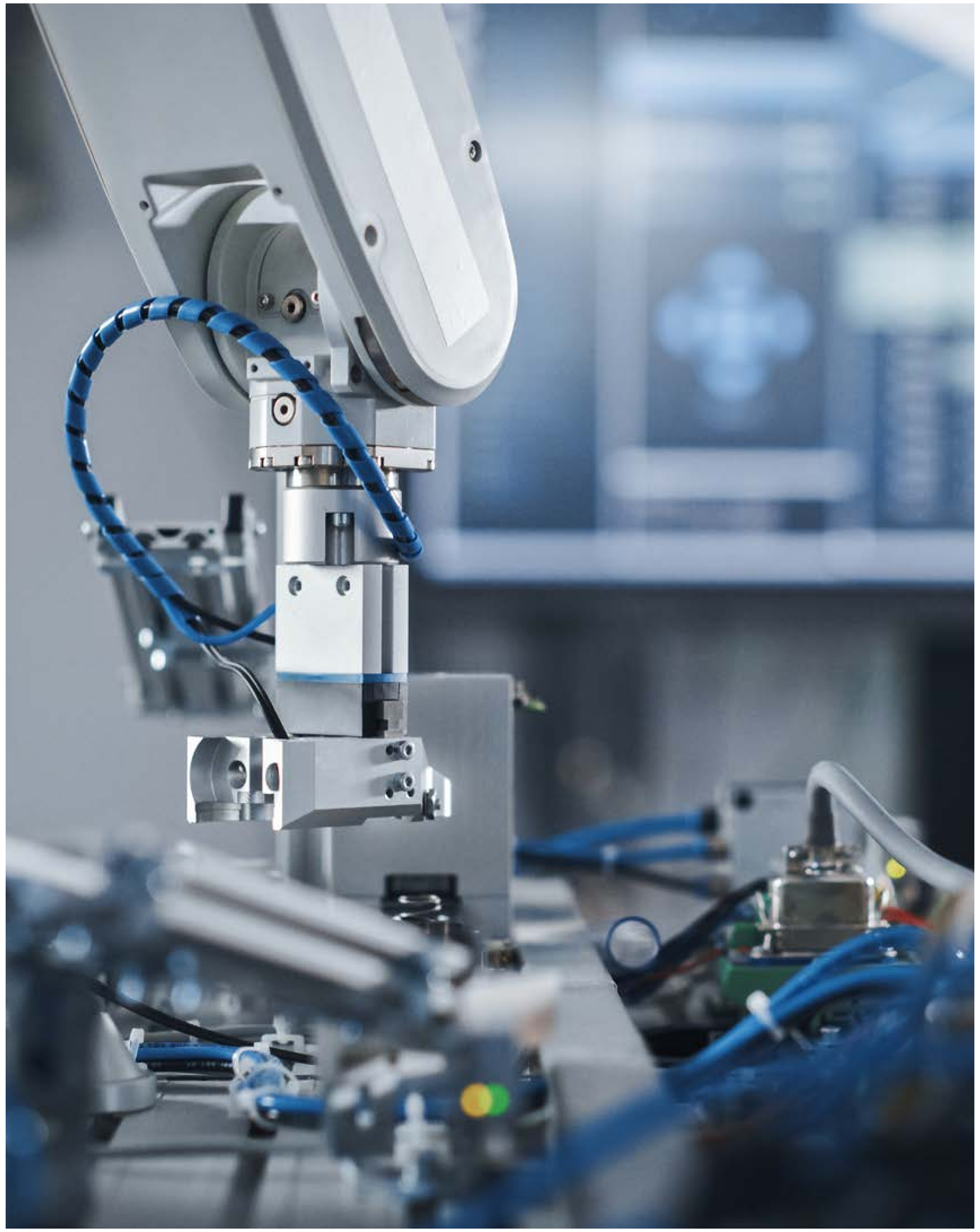
Relevant AI branches: All

Relevant business sectors: All

Potential financial impact: Depending on the nature and impact of the cyber attack, a brand's equity and value can be damaged and businesses can face consequences ranging from brand damage to fines and operational disruptions.

Example incidents:

- MULTIPLE INDUSTRIES
  In recent years, several large companies including Ikea, Yum! Brands, and T-Mobile have experienced data breaches where attackers used AI technology to expose employee and customer data[36].
- INFORMATION TECHNOLOGY
  Researchers are increasingly concerned about the role of LLMs like ChatGPT in cyberattacks because they can allow a wider range of malicious actors, including those with poor or no software development skills, to carry out complex cyberattacks. When prompted, ChatGPT is able to create working malware and generate sophisticated phishing emails[37].



Physical safety

AI is used to automate machinery that interacts with humans across a range of sectors (e.g. operating automated vehicles, industrial machinery, and performing surgery). By removing the risk of human error, AI can have a powerful impact on workplace safety in high-risk occupations and in patient safety in medical applications.

However, these applications where safety is dependent on AI, carry a risk of physical harm due to malfunctions or the poor functioning of the AI-controlled equipment. AI systems are in their infancy, so many vulnerabilities are still unknown and untested, and accidents may occur due to unexpected situations and environments as well as software and hardware failures[41].

Relevant AI branches: Robotics, ML

Relevant business sectors: Industrials, consumer discretionary, consumer staples, materials and healthcare

Potential financial impact: Companies whose AI use harms physical safety may face fines and/or criminal charges, litigation, or lose their social license to do business.

Example incidents:

- **CONSUMER DISCRETIONARY**
  The self-driving car company, Cruise LLC, was forced to recall 950 cars due to a software defect that caused a vehicle to run over a pedestrian even when the car 'slammed' on the brakes. The California Department of Motor Vehicles revoked the company's driverless testing permits[42].

Adverse generated content

GenAI works by mimicking input data to create new output data. When these associations are immature, malfunction, lack safeguards or are misused (accidentally or deliberately) they can generate several types of adverse content[43]:

- **Low quality** outputs and misrepresentation are common when general-purpose text and image generators attempt to generate complex, subject-matter specific or creative content.
- **Misinformation** occurs when LLMs perceive false connections, fabricating incorrect or nonsensical outputs that look plausible, known as AI hallucinations.
- **Disinformation** occurs when false information is deliberately spread through AI generated content such as fake reviews. **Deepfakes** are a specific form of disinformation where a person's voice or appearance is artificially simulated in video or audio, usually without consent.
- **Psychologically harmful, aggressive or obscene** outputs may be generated when chatbots and similar applications lack content safeguards (e.g. regarding swearing, pornography, or mental health).

Overreliance on GenAI, which results in these outputs, leads to an erosion of trust in the deployer and in AI more widely. Using low-quality generated content to make decisions without human verification may lead to physical, psychological or financial harm for the end user. As LLMs improve, easily detectable hallucinations and low-quality outputs decrease, which may lead to an erroneous assumption that outputs are becoming more truthful, despite the fact that they cannot distinguish between true and false information.

These outputs also contribute to wider **information pollution** – where information supply is contaminated with irrelevant, unsolicited, and low-value information. When such information is used again as inputs to AI generation, this leads to worsening **data pollution**.

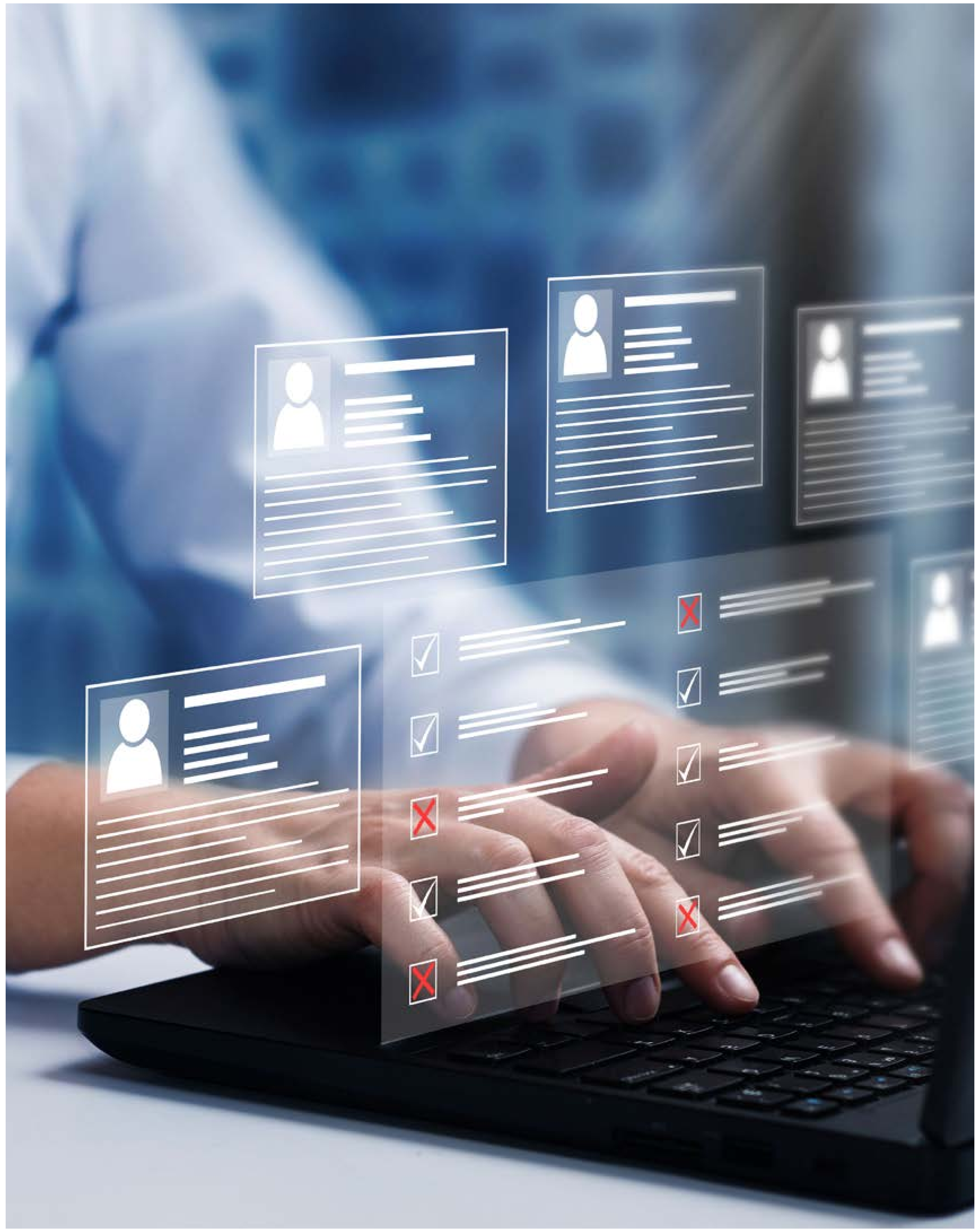
Relevant AI branches: GenAI

Relevant business sectors: All

Potential financial impact: Companies that produce adverse generated content may face a range of consequences depending on the level of harm, from brand damage, to litigation, fines and/or criminal charges.

Example incidents:

- **CONSUMER DISCRETIONARY**
  Netflix; Lego; drawing tablet company, Wacom; publisher, Bloomsbury; vendor, Calgary Farmers Market; and fashion brand, Selkie, have faced consumer backlash for using AI-generated art in their marketing, with detractors citing poorly rendered images and ethical issues[44,45,46,47,48].
- **IT, CONSUMER DISCRETIONARY**
  AI images of mushrooms on Google Images[49] and AI-generated foraging books by Amazon and other e-publishers[50] pose a significant health risk to consumers who (often unknowingly) use them as identification sources, leading to at least one case of poisoning[51].
- **FINANCIAL SERVICES**
  Voice deepfakes are used in financial fraud. In two known cases, fraudsters used deepfakes to pose as senior executives of the firm, convincing employees to transfer large sums (€220,000 and $25 million respectively) to false accounts[52,53].
- **IT**
  Microsoft shut down its Twitter (now X) chatbot after it became 'racist' and 'sexist' through unsupervised learning from other accounts within the same social media platform[54,55]. Two incidences of suicide have been linked to extended interactions with different AI chatbots[56,57].
- **IT**
  In 2025, Apple disabled a new iPhone update which hallucinated fake news alerts under real news providers' logos, after complaints from the BBC and National Union of Journalists, among others[58].
- **PROFESSIONAL SERVICES**
  In 2023, two US lawyers were fined $5,000 for submitting fake citations in a court filing. The citations were fabricated by ChatGPT in response to user prompts. ChatGPT further hallucinated case files to match the false citations[59,60].
- **HEALTHCARE**
  The US National Eating Disorders Association (NEDA) shut down its chatbot after it gave weight-loss advice to users seeking help for eating disorders[61].



Data privacy and security

Data privacy risks are not limited to AI but increase significantly in scale due to AI's vast data requirements. These requirements may motivate companies to collect sensitive data beyond what is deemed necessary.

The acquisition, transfer and usage of input data for AI systems is as yet unregulated and opaque. General-purpose models scrape data from the internet, and specific industries extensively utilise protected categories of customer data for AI inputs, such as financial, health and biometric data.

Data privacy and security laws like the EU or UK GDPR, which stipulate that only the minimal required data be used for any specific purpose, are in tension with the vast data processing requirements of AI systems, already resulting in several companies receiving sanctions[38]. Their implementation in instances of AI usage is not yet clearly defined.

Data privacy is closely linked to the risk of cyber attacks.

Relevant AI branches: ML

Relevant business sectors: All but particularly financial services, healthcare and IT

Potential financial impact: Companies violating data privacy laws can face severe fines.

Example incidents:

- IT
  Clearview AI, a facial recognition platform, was sued in the UK, Netherlands, Italy and France for unlawfully storing images of faces scraped from social media websites.

  The images allowed Clearview to identify their subject's personal information. Some of Clearview's key clients, such as Interpol and the NYPD, distanced themselves from the company and claimed to have stopped using its products[39].
- FINANCIAL SERVICES
  In Hungary, the Data Protection Authority fined a bank almost €700,000 for using AI voice analytics on customers in its call centres, citing a lack of consent or 'legal basis', inadequate security safeguards, and a failure to notify data subjects of their rights[40].

Intellectual property infringement

AI-generated content must be trained on existing content. AI texts, images, videos, and audio therefore have varying similarity to existing content and may be purposefully or accidentally directed to copy existing content, raising concerns around the infringement of intellectual property (IP) rights. Several aspects of risk exist:

- Companies in media, publishing, advertising, film and other creative industries face infringement of their IP rights. This may be identified in the input stage where proprietary content is used as training data and in the output stage where AI content is generated which bears strong resemblance to existing IPs.
- Businesses using AI generated content for purposes such as marketing may be at risk of litigation from parties who believe their content has been copied.
- Whether generated content itself can and should be protected by IP laws is currently unresolved and subject to ongoing cases and policy discussions.

As the technology matures, protected properties such as patents and product designs may face similar IP infringement concerns as text, image and multimedia content, exposing a wider range of sectors and business functions to IP risks.

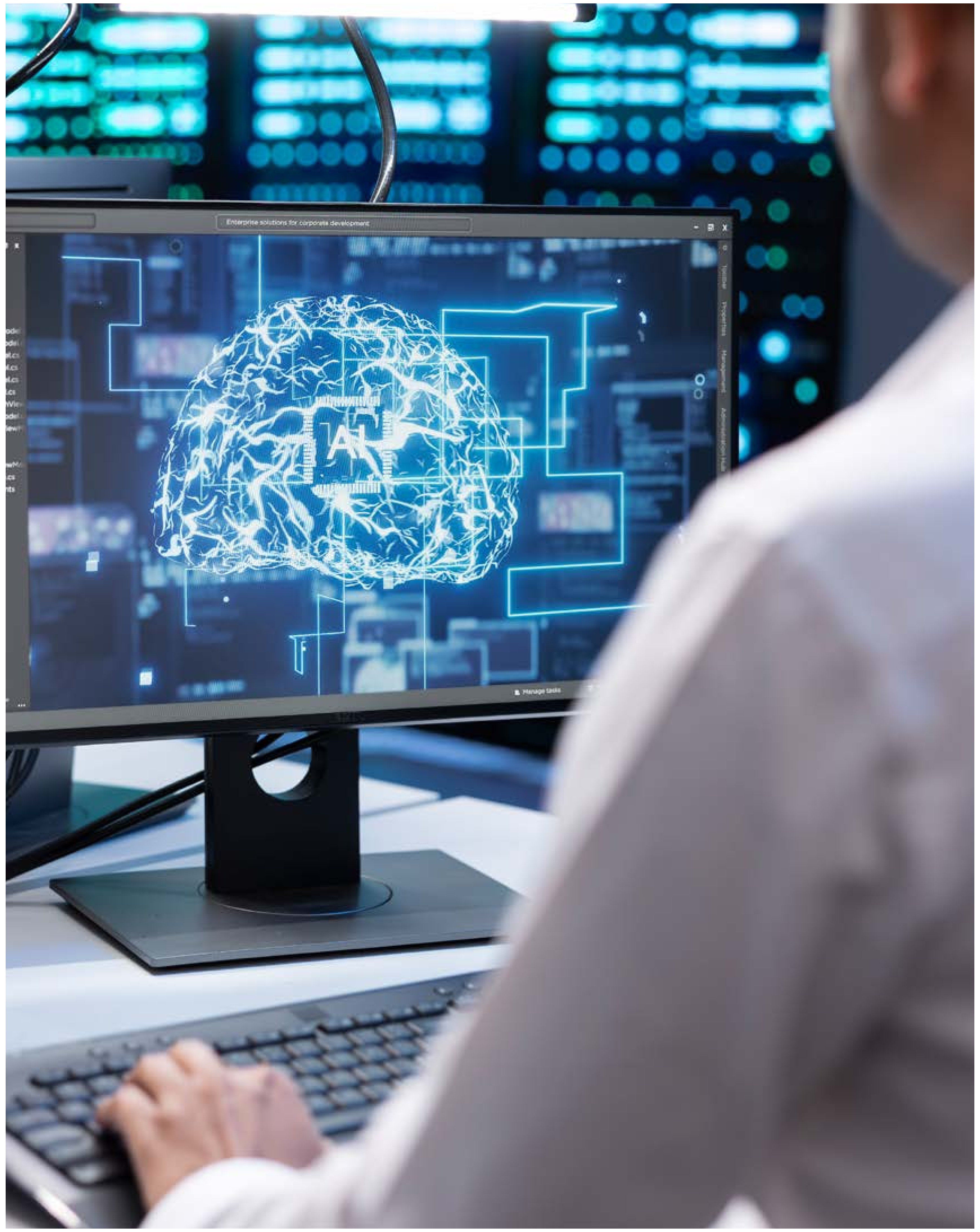
Relevant AI branches: GenAI

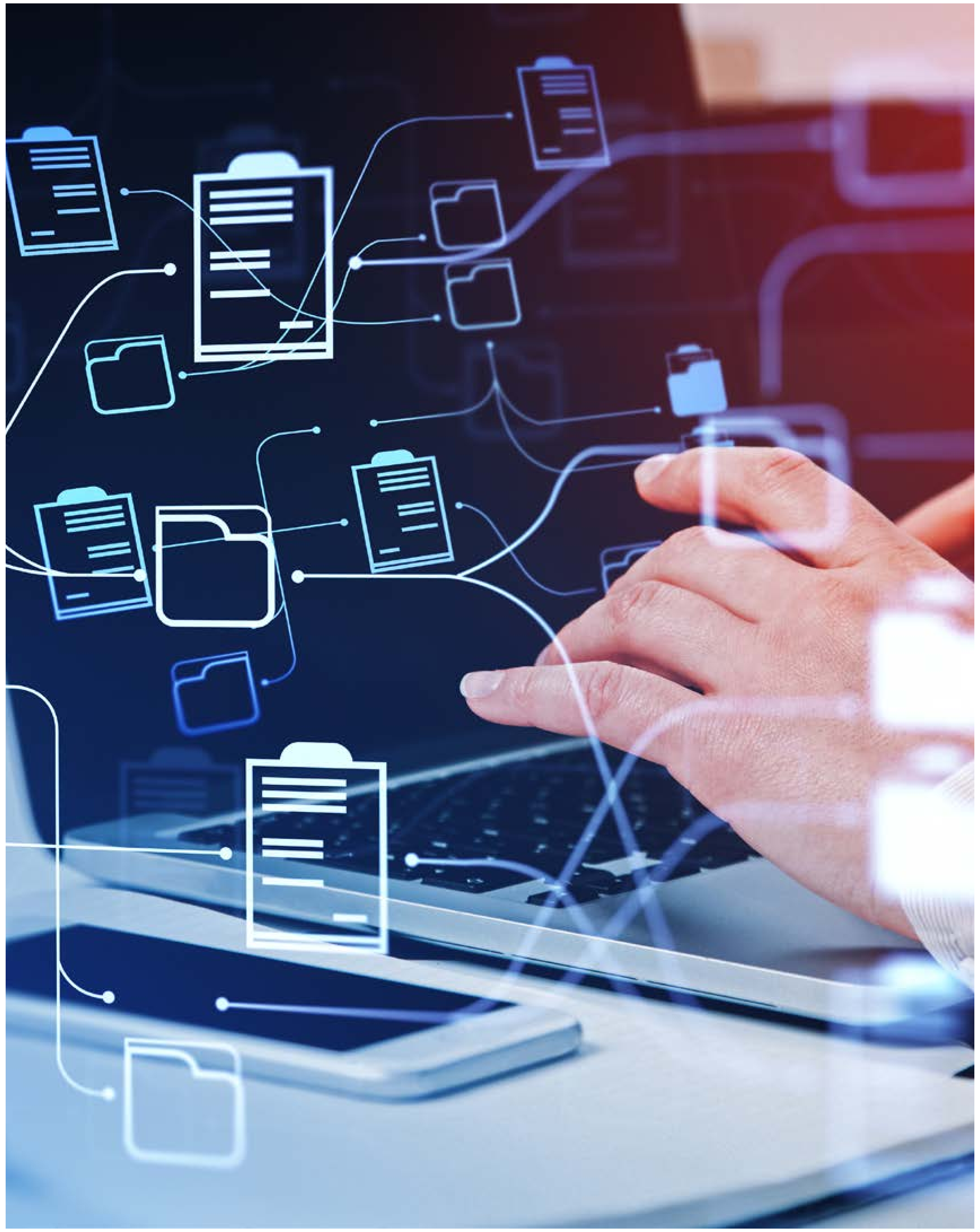
Relevant business sectors: All, particularly consumer discretionary

Potential financial impact: Companies infringing IP rights may face fines and litigation, while those whose IP is infringed may face brand dilution leading to a decreasing customer base.

Example incidents:

- **CONSUMER DISCRETIONARY**
  A film production company is suing Tesla for usingan AI-generated imitation of its film scenes to promote its vehicles[65].
- **CONSUMER DISCRETIONARY, IT**
  There are now numerous cases of legal action by news publishers and book authors whose copyrighted works have been included in training data for LLMs. The defendants have included Nvidia, Databricks, Anthropic, Perplexity AI, OpenAI, and Microsoft[66,67]. The complaints generally cite direct or vicarious copyright infringement, unfair competition, and trademark dilution.
- **CONSUMER DISCRETIONARY, IT**
  Major recording companies have filed lawsuits alleging direct or vicarious copyright infringement against the companies behind apps, Suno and Udio, which allow users to generate digital music files based on text prompts[68].





Transparency and explainability

Transparency ensures users are able to understand how AI processes work, while explainability refers to whether it's clear how a specific AI outcome or decision was arrived at.

As DL-based systems become increasingly complex, it's becoming more difficult to identify the origin of their outputs. If AI outputs can't be traced and explained, their accuracy or effectiveness can't be easily evaluated.

Black box AI models become an issue when they're used to inform decisions and where decision-makers and affected parties don't have a clear understanding of the inputs and processes.

Organisations may face scrutiny if they use proprietary models without understanding and disclosing their processes around the data used as inputs, how this data is interpreted to create outputs, and tests for biases.

Transparency risks are closely related to accountability, data privacy and cultural bias.

Relevant AI branches: All, particularly ML

Relevant business sectors: All

Potential financial impact: Companies that are deemed to over-use AI and/or lack transparency on the extent of AI use in key decisions may face litigation and reputational consequences.

Example incidents:

- **HEALTHCARE**
  UnitedHealthcare and Humana – two major US health insurance providers, are facing a class action lawsuit regarding their use of an AI algorithm to make insurance decisions.

  It's alleged that they've denied care to elderly patients, prematurely terminated cover and pressurised clinicians to conform to the algorithm over their own expertise. When insurance denials are appealed, most are reversed, suggesting the AI decisions were not accurate nor robust enough to be defended[62,63].
- **PUBLIC SECTOR**
  In multiple US states, judges use the COMPAS system, which scores perpetrators on their likelihood of reoffending, to aid sentencing decisions. Research has shown the system favours jailing over release and displays age and race-related biases. Neither the public, jury nor judge are privy to the evidence COMPAS uses to generate its outputs[64].

Labour and employment

AI has the potential to transform the nature of labour across the whole economy, although estimates of employment gains/losses and wage growth/stagnation remain highly speculative.

It's been estimated that up to 8 million UK jobs may be at risk and 11% of tasks are already exposed to the 'first wave' of automation[69]. Lower paid 'routine' cognitive and organisational tasks are at highest risk, with a disproportionate effect on women and youth[70]. GenAI increasingly presents a risk to traditionally 'safe' jobs in communications, software engineering and creative industries. Employees are impacted by job replacement or the fear of it. The replacement of creative jobs is closely related to IP risks.

AI-related workplace issues include:

- **AI deployment:** AI use may cause productivity and retention risks resulting from a lack of familiarity, poor training, perceived biases in recruitment, and poor working conditions such as extensive surveillance and monitoring[71,72]. Large-scale pushback from workers can present a significant risk to industries harnessing AI for management and value creation.
- **AI development:** AI models still require vast amounts of human input in areas such as data labelling and content moderation – often performed by low-paid workers in strict conditions[73].

Relevant AI branches: All, particularly GenAI

Relevant business sectors: All, particularly IT and consumer discretionary

Potential financial impact: While labour changes can lower costs, worker fear and resistance to AI could cause operational disruption leading to loss of revenue and reputation. Companies using AI for recruitment and termination may face litigation.

Example incidents:

- IT
  Open AI and Tesla are amongst the companies that have been critiqued for basing AI developments on the intensive workloads and automated surveillance of data labellers, content moderators and warehouse workers[74,75].
- MULTIPLE INDUSTRIES
  Survey data shows that two thirds of US and Canadian contact centre employees agree automated monitoring made work more stressful[76]. A survey of 4,000 employees across the globe found that only half welcome AI in the workplace and 42% believe their company does not have a clear understanding of which systems to fully automate. In total, 1 in 4 employees aren't confident their organisation prioritises employee interests and responsible AI and implementation[77].
- CONSUMER DISCRETIONARY
  Labour unions including the Association of Australian Voice Actors[78], Screen Actors Guild, Writers Guild of America[79], and Animators Guild[80] have raised concerns about the risks GenAI presents to jobs in these industries. Created content (e.g. books, artworks) and individual likenesses (e.g. voice, appearance) are increasingly used as AI inputs to new products, raising ethical concerns around workers' consent and knowledge, compensation and job losses. Following strike action, the Writers Guild won protections requiring studios to disclose whether material given to writers was developed with AI[81].
- CONSUMER DISCRETIONARY
  Uber, Estee Lauder, and Workday Inc (among others) have been sued for their use of AI in hiring and in making redundancy decisions[82,83,84].

We think the theoretical risk examples on pages 12-19 clearly articulate why investors must deepen their dialogue with portfolio companies on approaches to AI, while also beginning to recognise this as a system-wide issue (i.e. having cross-sector and cross-jurisdictional impacts).

This dialogue should take place with companies far outside the IT sector (or AI developers), where such engagement has understandably been concentrated. As this report explains later, a system-wide stewardship approach to AI should also include engagement with policymakers, regulators and standard-setters, to help achieve the necessary system-wide framework and solutions.

HOW CAN INVESTORS ASSESS THE SIGNIFICANCE OF AI TO THEIR PORTFOLIO COMPANIES?

This section expands on our understanding of the system-wide financial materiality of AI by detailing the essential elements of risk assessment that investors should consider for effective AI governance.

AI is used in nearly all sectors, with 72% of companies having adopted it in at least one business function during 2024[85]. Therefore, investors must identify where AI is, or may in the near future be, most significant for the companies in their portfolios. This identification is crucial for prioritising their stewardship efforts effectively.

The level of risk (and opportunity) a company faces varies based on its role in the AI value chain, its operational dependency on AI, and how its sector uses AI. When conducting a high-level risk assessment, we recommend asking the following questions:

1. Is the company an AI developer, deployer, or both?
2. How significantly is AI being used within the company?
3. How does the company's sector use AI?

The distinction between AI developers and deployers

Developers of AI are the primary targets of existing safety measures, and currently face more direct scrutiny than deployers[86]. While investors may initially direct their attention to developers, companies in both roles face risks.

A developer cannot control the use of their AI system by a deployer and a deployer might not be held responsible for biases embedded in the system during the design stage by the developer[87]. This is beginning to be recognised by regulators, with the EU AI Act specifying responsibilities for deployers of high risk AI systems.

The distinction between the role of a developer and that of a deployer is being increasingly blurred; indeed a company can be both a developer and a deployer. However, where the company is one or the other, they could potentially be held liable for the actions of the other despite those actions being out of their control.

Developers of AI design, code, and test models and bring them to market. These include prominent IT companies such as OpenAI and Alphabet, but also businesses across other sectors including finance, manufacturing and biotechnology.

Deployers are third-party users, i.e. entities that use an AI system in a professional scope but did not create that system.

Deployers that customise AI systems, incorporate extensive proprietary data, and design derivative services may take on a similar role to developers[g]. Companies may be both developers and deployers.

For example, Microsoft develops LLMs and deploys them to end users via its Copilot virtual assistant.

[g] Hewson (2024). The roles of the provider and deployer in AI systems and models. The roles of the provider and deployer in AI systems and models.

The significance of AI at company level

The more a company relies on AI, the greater the related risks. High dependency on AI means that incidents can lead to larger financial, operational, legal or reputational consequences. Therefore, categorising companies by AI significance provides a practical way to allow investors to prioritise where the risk will be most material.

We note however that social, legal and technical understanding of AI impacts are constantly evolving, and risks can occur at any level of significance. The EU AI Act, for example, does not distinguish between companies deploying AI in core functions or supporting business functions; although it does apply differentiated responsibilities to high-risk uses[88].

Figure 3: **Level of AI significance**

Level of AI significance	Characteristics	Deployer and/or developer
Low AI significance	• Use of off-the-shelf AI systems • Used in single/few ancillary business functions • No integration into high-risk processes	Deployers only
Medium AI significance	• Use of off-the-shelf AI systems with some customisation • Used in multiple ancillary business functions, e.g. marketing, administration • No integration into high-risk processes	Deployers only
High AI significance	• Development of AI systems • Extensive customisation of AI systems • Integration into core business functions • Integration into any high-risk processes, e.g. where it impacts physical safety of assets and people, large financial transactions, or large amounts of protected data is being processed	Deployers and developers

The use of AI at sector level

High AI significance exists in certain sectors, including IT, healthcare, and finance. This is due to the extensive use of sensitive data in AI applications, alongside the heightened impact of decisions made based upon AI-driven insights. For other sectors, AI risks shouldn't be generalised as dependence on AI varies considerably between companies, meaning all sectors could have companies with high, medium or low AI significance.

While we don't therefore map AI risk directly to sectors, understanding how peer companies in a sector are already using AI can be helpful in determining the potential future trajectory of a given company's own approach to AI. Figure 4 (right) identifies 11 key sectors and demonstrates how they utilise AI in industry-specific ways[89].

Once investors consider and answer the questions we pose on page 20 of this report as part of their portfolio-wide risk assessment, they can use this to help them understand where and how they might meaningfully deploy their usual stewardship tools. This is the subject of the next section of our report.

⚠ = High AI significance

Figure 4: **Current AI usage within sectors**

Sector	AI usage
1. IT ⚠	AI products are largely developed and deployed to other sectors and end-users by IT companies. Industries across the sector, such as semiconductor manufacturing, hardware and electronics, and software engineering are necessary for AI development and deployment.
2. Healthcare ⚠	The use of AI includes diagnostic imaging for disease identification, predictive analysis for health insurance premiums, and robotics for automated surgery and equipment. AI in healthcare was valued at USD 19.27 billion in 2023, predicted to reach USD 613.8 billion by 2034[90].
3. Financial services ⚠	AI use includes fraud detection and pricing of insurance and mortgages. Investment managers may use AI for investment analysis and to optimise portfolios and asset allocations. In 2020, 85% or financial organisations used AI and 77% expected AI to possess high importance to their business within two years[91].
4. Industrials	As well as predictive maintenance of assets and equipment, AI may be used to enhance risk analysis, testing, and other safety protocols. It may be employed to optimise route planning for sectors such as rail, aviation, and logistics.
5. Consumer staples	In 2024, 72% of retail decision makers in the US felt 'ready' to deploy GenAI technology[92]. AI is already used for forecasting in supply chains, inventory and distribution management, and automated production lines (e.g. for quality checks, handling and packaging).
6. Consumer discretionary	AI is used to innovate products and services across a range of discretionary goods, e.g. in automotives, retail, and leisure. AI creates personalised shopping experiences and smart product recommendations. AI-based consumer products are increasing, e.g. Tesla Autopilot, Adobe Firefly.
7. Materials	AI may be used to optimise production processes through real-time adjustment of manufacturing conditions, quality control and detection of defects in products and equipment.
8. Utilities	AI can improve smart-grid management by forecasting electricity demand, managing grid complexity and predictive equipment maintenance. In 2023, a third of utility and energy companies globally were piloting GenAI[93], increasing to 74% in 2024[94].
9. Energy	Energy companies may use AI for predictive maintenance and adjustment of equipment, such as oil rigs, and increasing operational efficiency, e.g. adjusting wind turbines according to weather conditions. AI can also help forecast energy prices.
10. Communication services	In 2023, 90% of telecoms companies used AI, with 48% piloting and 41% actively deploying AI. The uses of AI included predictive analysis of network usage and automated network management, e.g. traffic routing and capacity planning to avoid outages.
11. Real estate	Potential uses of AI include algorithmic property valuations and property management uses such as predictive and scheduled maintenance of HVAC systems.

WHAT CAN INVESTORS DO TO MANAGE AI RISKS?

The pace and scale of AI adoption continues to outstrip the development of regulatory frameworks.

Therefore, investors have a critical role to play in addressing the risks we outline in this report – particularly where portfolio companies are deeply embedded in the AI value chain, reliant on AI systems, or operating in sectors where AI use is rapidly evolving.

Managing these risks effectively requires collective action. We encourage investors to participate in dialogue with companies who are at the forefront of AI development and deployment.

We also encourage investors to proactively feed into the emerging policy and regulatory discussion on the management of AI risks, and the effective harnessing of AI opportunities. Our sense is that the investor perspective is missing from these policy debates.

In the following section, we summarise current investor activity and set out our proposed approach to engagement and policy advocacy on this financially material issue.

Investor activity

A 2024 survey found that two thirds of 1,130 institutional investors believe the social risks of AI are already material and over half were concerned about its environmental risks[96]. Investors' attention towards AI increased in 2024, driven by accelerating technological advancement, increasing commercial deployment, and a growing recognition of the material social and environmental implications associated with its use. Investors are engaging engaging in four main ways:

1. Publishing expectations
2. Engaging with companies
3. Submitting and voting on shareholder resolutions
4. Participating in collaborative initiatives.

We could not, as we outline later, find much evidence of concrete and proactive investor participation in policy debates.

What does Railpen do on AI?

Our Global Voting Policy

Our 2025 Global Voting Policy builds on the foundations we set in 2024, when we first introduced AI as a dedicated focus, by strengthening expectations around its responsible use, recognising both its transformative potential and the expanding range of associated risks. We expect companies developing or deploying AI to demonstrate accountability across the AI value chain, with actions proportionate to their risk exposure, business model, and potential impact. This includes clear board oversight, robust risk management, and transparency.

Where these expectations are not met, and there is evidence of egregious social or environmental harm and inadequate governance, Railpen may vote against the director responsible for oversight. We may also support shareholder resolutions addressing AI-related reporting, board accountability, human rights, misinformation, and workforce implications. Our expectations will continue to evolve with the technological and regulatory landscape.

Collective initiatives

In 2019, Railpen joined a collaborative investor initiative led by Royal London Asset Management to address the systemic risks associated with cybersecurity. This Cybersecurity Coalition engages with portfolio companies to promote stronger cyber governance. In 2025/26, the coalition's engagement will focus on the intersection of AI and cybersecurity, specifically, the risks and opportunities AI presents and the practices companies can adopt to manage them effectively. Through this engagement, we aim to uncover best practices and encourage greater transparency in how investee companies approach AI-related cyber risks.

Railpen is also part of the Big Tech and Human Rights investor initiative, a collaborative engagement launched in March 2023. This initiative brings together institutional investors to address systemic risks linked to the human rights impacts of major technology companies. It aims to drive positive change and strengthen the integration of human rights considerations across the digital ecosystem. Given the close interdependencies between AI and Big Tech, the engagement also includes focused dialogue on AI governance and accountability.

Direct engagement and policy advocacy

Railpen's involvement in the Cybersecurity Coalition led by Royal London Asset Management sees us undertaking direct engagement on the intersection of cybersecurity and AI. In addition, AI is being incorporated into broader engagements with our most material holdings.

Following the publication of this report, Railpen will continue to evolve its stewardship approach by developing a dedicated AI strategy. As we are at an early stage in shaping a coherent approach to responsible AI stewardship, we have not yet undertaken policy engagement in this area, but we recognise the opportunity and intend to pursue it as our thinking matures.

Real Assets: Our approach to AI-related investment opportunities

Railpen's Real Assets strategy focuses on investing in key areas of growth in the UK economy, spanning real estate and infrastructure. The team is actively exploring the long-term opportunities presented by AI, and we recognise the transformative potential of AI to shape the world our members will retire into.

Our commitment to investing in the Oxford-Cambridge arc demonstrates how we are supporting a leading hub for AI innovation. Cambridge is home to a world-class university, a thriving STEM ecosystem, and a globally recognised AI research community. Our strategy focuses on creating an innovation cluster of high-quality, sustainable real estate, comprising research space, offices, residential and community space. These developments place us at the heart of a region poised for significant growth in science and technology, with the potential to deliver long-term value for our members.

Within infrastructure, we see AI as a catalyst for growth in the UK's digital economy. AI is already playing a role in optimising revenue trading within our utility-scale battery storage assets. We are also exploring direct investment opportunities in infrastructure needed to support AI technologies, including data centres. These assets can exhibit core infrastructure characteristics, but we understand that the market faces challenges around power, planning, and connectivity. We are therefore engaging with the UK Government to find ways to deliver the required solutions to these challenges.

Our decision to allocate capital to AI-related opportunities reflects a considered assessment of its long-term financial characteristics, weighed against material risks. We are mindful of the specific risks associated with AI – including some of those outlined in this report – and expect each asset to present a credible plan for addressing them. We conduct an ESG risk assessment early in the investment process, including environmental considerations such as access to co-located energy sources and systems to minimise water intensity. These challenges are not a deterrent,but rather an opportunity to deploy capital responsibly and drive positive change as an active asset owner.

Fundamental Equities: How Railpen's growth-oriented active equity portfolio considers AI tailwinds

Our growth-oriented, bottom-up equity portfolio targets structural growth trends, with AI emerging as one of this decade's most powerful themes. As part of our research and investment process, and working together with colleagues across the business, we systematically evaluate how our investments might benefit from or be threatened by AI-driven disruption.

Key AI themes in our portfolio

- **Enabling AI infrastructure:** Portfolio companies like Taiwan Semiconductor Manufacturing Company are important for the AI transition, supplying advanced chips essential for AI applications. Meanwhile, Microsoft's Azure cloud and partnerships, as well as Amazon's AWS and Equinix's data centre offerings, form the backbone for other firms' and sectors' expanding AI capabilities.
- **Supporting AI-driven business transformation**: Large enterprise software firms such as ServiceNow are harnessing agentic AI - systems that can achieve a specific goal with limited human intervention - to help clients automate and streamline operations, helping to drive greater efficiency and support digital transformation.
- **Enhancing security and data analytics**: Firms like Palo Alto are leveraging AI to better protect enterprise clients from evolving cyber threats, while data analytics companies like RELX are increasingly using AI to automate research, compliance, and fraud detection.
- **Monetising and integrating AI**: Microsoft leads in commercialising generative AI through products like Copilot, and is strategically positioned via its partnership with OpenAI. Similarly, Meta's investment in open source large language models (like Llama) strengthens its core advertising business and differentiates it from smaller peers.
- **Meeting AI-driven power demand:** Energy companies such as NextEra and utilities like PSEG are expanding capacity, particularly in renewables and nuclear, to meet the surging electricity needs of growing data centre infrastructure.

While our portfolio is structured to benefit from the growing integration of AI across industries, we recognise the importance of good governance at these companies to help them effectively harness the significant opportunities AI presents for their business models, as well as manage any associated risks.

Emerging investor stewardship initiatives and activities on AI

The information on pages 27 and 28 of this report summarises our understanding of the current landscape of emerging investor stewardship initiatives and activities on AI. We find that the focus of investors' activities has primarily been limited to large technology companies, and to company-focused as opposed to policy engagement.

We then suggest an approach to help investors put into practice what are often, at this stage, theoretical 'principles' for AI, through our proposed AIGF, alongside good practice case studies on the governance disclosure that will help investors better understand company approaches, suggested questions for company engagement, and what a meaningful approach to policy engagement on AI might look like.

Publishing investor expectations

Description: Investors are beginning to publish their expectations of governance, strategy, risk management and transparency for AI for portfolio companies.

Key examples:

- Hermes Investment Management published its Investors' Expectations on Responsible Artificial Intelligence And Data Governance in 2019[97]. The report describes AI's relevance to long-term investors through its application in key sectors, its responsible AI principles, salient social risks (i.e. AI-specific risks) and outlines a governance analytical framework for companies.
- LGIM's 2023 Autumn CIO update[98] outlines its governance, risk management and transparency expectations of companies, with further disclosure on how it is voting according to these expectations[99,100].

Engagement with companies

Description: Some investors have conducted engagement with companies that develop or use AI extensively, particularly ahead of key AI-related resolutions or to provide feedback on their AI-related publications.

Key examples:

- Abrdn[101] has met with Apple regarding its publication of ethical guidelines related to AI and also with Meta to express concern around its due diligence standards around AI advertising. It also provided feedback to Microsoft on how to evolve its AI Transparency Report.
- Federated Hermes EOS regularly engages with GSK on behalf of Brunel Pension Partnership. In 2024, EOS provided GSK feedback on its draft Responsible Use of AI Policy, highlighting oversight and reporting procedures. GSK incorporated this into its final policy[102].

Shareholder resolutions and voting

Description: AI-related shareholder resolutions increased significantly in 2024, focusing on large tech companies and entertainment companies. None to date have received majority backing, but there is a trend of increasing vote share.

A global consultancy firm analysed 23 AI-related proposals in the 2024 AGM season, finding they largely concerned transparency on companies' AI usage, policies and processes, with some specific risks raised, such as disinformation/misinformation and implications of AI use on the workforce[103]. Alongside the rise in shareholder resolutions, investors are also beginning to integrate lines in their voting policies relating to the responsible use of AI

Key examples:

- A proposal raising concerns that Microsoft may be using unethical or illegal data sources to train its GenAI models was supported by 36% of investors[104], including LGIM and Storebrand Asset Management[105].
- A proposal filed by the American Federation of Labor and Congress of Industrial Organizations calling for transparency on the use of AI by Netflix garnered 43% of the vote[106].
- Less common sectors receiving AI-related proposals for the first time in 2024 included hospitality (Chipotle) and healthcare (UnitedHealthcare Group)[107].
- Railpen's Global Voting Policy includes a dedicated section on Responsible AI (see page 24).

Collaborative initiatives

Description: Investor expertise has a role to play in advancing the implementation of Responsible AI through multi-stakeholder initiatives.

Key examples:

- In 2024, CPP Investments partnered with the World Economic Forum to develop a Responsible AI Playbook for Investors. The report establishes investors' role in promoting responsible AI and provides practical tools to develop their approach[108].
- The Collective Impact Coalition, conceived by the World Benchmarking Alliance joined by 33 investors, engaged 28 tech companies to advance ethical AI policies and practices[109]. The project found that governance mechanisms, impact assessments, and due diligence around buying and selling AI products are unclear. However, companies are increasingly consolidating their AI policies, with a majority referencing human rights standards.
- Alphinity Investment Management collaborated with the Australian Government's Commonwealth Scientific and Industrial Research Organisation to publish a governance framework with an escalating approach for investors to manage AI risks[110].
- The Cybersecurity Coalition led by Royal London Asset Management (of which Railpen is a member) is now incorporating AI into its dialogue on cybersecurity risk and resilience, as part of its evolving engagement strategy.

Responsible AI principles – our summary

To date, many organisations across industry, government and civil society have published values-based AI principles, with the United Nations Educational, Scientific and Cultural Organisation (UNESCO) and the Organisation for Economic Co-operation and Development (OECD) AI principles as early reference points[111,112].

Further sources include the Partnership on AI[113], and the Responsible AI Institute[114]. Some specifically target investors, such as the CFA Institute's guidance on Ethics and Artificial Intelligence in Investment Management[115], the International Corporate Governance Network's (ICGN) engagement guide for investors[116], and the Responsible AI Institute's Guiding Framework for Responsible AI Integration into ESG Paradigms[117].

However, far less work has been done on how to incorporate and action these principles in specific industries and organisations.

The table (right) aggregates common AI principles, adapted from aforementioned guidance and multiple corporate responsible AI policies. Each principle addresses at least one key AI risk and they are each relevant to most businesses developing and deploying AI.

Principle	Description
Validity and reliability	AI inputs and outputs should be up-to-date, valid, truthful and reflect reality. There should be human processes to regularly check AI for false or poor-quality outputs, disinformation and misinformation.
Safety	Human safety should be prioritised over revenue and technological development. AI systems should be regularly monitored to ensure they function appropriately and do not pose unreasonable safety risks. Mechanisms should be in place to override, repair, and/or decommission machines exhibiting undesirable behaviour.
Fairness and inclusion	AI algorithms and the adoption of AI should not replicate societal biases or cause disproportionate systematic harm to certain communities or groups.
Security and resilience	AI usage should be protected from outside interference and abnormalities in input or output, whether malicious or unintentional, that could compromise the security or resilience of the system.
Privacy	AI systems should process personal information in a secure manner and in accordance with data protection laws and principles (e.g. GDPR). AI surveillance should not infringe labour rights and human rights.
Explainability & transparency	End-users (employees and customers) should understand how the system works, know when they are using AI, and have clear, accessible information on its sources, benefits and risks, and how to interpret outputs. Datasets, processes and decisions made during the AI system lifecycle should be traceable.
Accountability	Humans should be accountable at the highest level of decision-making. Clear governance structures for the development, deployment and outcomes of AI systems should be implemented, including proactive monitoring and management of negative outcomes. Users should have a mutually agreed system of accountability with third parties.
Environmental sustainability	The environmental impacts (e.g. energy and water consumption) used during the AI lifecycle should be measured. These impacts should be weighed against the user's environmental policies and mitigated where appropriate.
Adaptation and continuous improvement	Users should recognise that AI is a fast-moving technology and should ensure monitoring, testing, and audit mechanisms stay up-to-date as AI evolves. All end users should be regularly trained to keep up with changes to AI processes and on the wider landscape and risks (technical, regulatory, legal, and social).

Moving from theory to practice

A guiding concept of recent investor activity in both individual and collective engagements has been something the industry generally calls 'responsible AI'. This concept involves the design, deployment, running and monitoring of AI systems in ways that consider broader societal impacts, empower organisations and people, and helps ensure equity for all stakeholders[118]. Responsible AI is guided by a set of principles that promote the safe, trustworthy and ethical use of AI.

These principles on are important and useful, and we commend the efforts of policymakers, investors and businesses in leading the way. However, we believe more should be done to help investors – especially investment stewardship practitioners – to put these principles into practice in their work with companies and other key stakeholders.

To move from theory to practice we, we believe the following are needed:

- An **AI Governance Framework (AIGF)** – to help investors understand companies' preparedness regarding AI risk (and opportunity).
- **Guidance on AI disclosures** – why this matters and what good looks like; what kinds of questions investors can ask their portfolio companies where disclosure falls short.
- **Advice on AI policy advocacy** – covering major developments and how investors can use public policy as a system-wide stewardship tool.

Our proposed AI Governance Framework

To help investors assess companies' approaches to risk management, we have developed a AIGF that moves the responsible AI principles described on page 29 from theory to practice. Although the long-term capabilities and associated risks of AI remain largely unknown, the framework on pages 31-33 can allow investors to understand companies' preparedness for these uncertainties as well as the steps that company may need to take to manage the risks and harness the opportunities.

Extending the focus beyond large technology companies, this framework is designed for a broad range of AI developers and deployers. Companies with medium and high AI significance should be assessed against these guidelines, with additional expectations required of companies with high significance (Figure 3 defines low, medium and high AI significance).

We recognise that companies may not yet be in a position to meet all expectations set out in the AIGF, as implementation will depend on their organisational maturity. However, we would expect to see a phased and deliberate approach to integrating the framework's pillars over time, as companies advance in their AI governance journey.

Pillar	Indicator	Core expectations for companies with medium and high significance	Additional expectations for companies with high AI significance
Governance	1. Oversight of AI	• The company has senior level oversight of AI risks • There is evidence that AI issues are discussed at the board and/or senior leadership level • There is board and/or senior leadership level training on AI usage and risks	• The company has board-level oversight of AI risks • There is an individual on the board with AI expertise
	2. Management of AI	• The company reports to the board on implementation of AI systems and processes through annual reviews • The company has individuals with management responsibility for AI deployment • The company implements training to build staff capacity on AI-related risks, principles and responsibilities	• There are dedicated AI experts within the company • There is a committee with a responsibility to coordinate on AI use across the company • The company allocates a dedicated budget to the management of AI risks
	3. AI-related policies	• The company maintains a standalone or integrated policy on AI use • The policy aligns broadly with recognised principles of responsible AI use • The policy is publicly available and accessible and is widely communicated within the company • The policy identifies specific AI risks to the business • The scope of the policy is clearly defined, and includes its own operations, supply chain, business relationships, customers, and other stakeholders • The policy states who is responsible for its governance and implementation • The policy outlines how it impacts and relates to other policies (e.g. sustainability, code of conduct, human resources, data processing, privacy, human rights, business ethics) • The policy explicitly refers to data management covering data ownership, sovereignty, processing and privacy • The policy is developed with relevant internal and/or external expertise	• The policy states how the company operationalises relevant principles of responsible AI use and has a ‘comply or explain’ approach to others

Pillar	Indicator	Core expectations for companies with medium and high significance	Additional expectations for companies with high AI significance
Strategy	4. AI relevance to business strategy	• The company has assessed how AI impacts its business model • The company can describe how AI is a strategic issue for the business	• AI expertise informs the company's strategic reviews
Risk management	5. Identification of AI risks and impacts	• The company has conducted an assessment to identify AI risks and impacts across its operations • The company has assessed how AI intersects with other issues in its risk register • Through due diligence, the company has assessed its AI suppliers and other suppliers' usage of AI and identified high-risk areas • The company has identified the objective using of a particular AI system and conducted due diligence to ensure that the system can deliver that objective • The company frequently updates its understanding of AI risk and how this is factored into risk management processes	• The company has a centralised system to monitor its AI usage • AI risk identification is reviewed through auditing processes (internal and external)
	6. Management of AI risks and impacts	• The company has developed an action plan around its salient AI risks, including controls around input data manipulation and data loss prevention • The company takes measures to prevent and mitigate specific AI risks in its operations and business relationships	• AI risk management is reviewed through auditing processes (internal and external) or through alignment to risk frameworks such as NIST AI 100-1

Pillar	Indicator	Core expectations for companies with medium and high significance	Additional expectations for companies with high AI significance
Risk management	7. Stakeholder engagement	• The company has a policy or guidance on stakeholder engagement regarding AI • The company updates all stakeholders on its AI usage in a clear and accessible manner • The company declares its AI usage to all end-users (e.g. employees, contractors, customers, job candidates) • The company collects feedback on its use of AI systems from end-users and experts	• The company uses data from stakeholder engagement to inform its understanding of its AI risks and impacts • The company describes its use of specific AI systems, including risks and benefits of use, in a clear and accessible manner to end-users (e.g. employees, contractors, customers) • The company has grievance channels for end-users to address malfunctions, complaints and adverse impacts of AI systems and ensures they are accessible • The company engages with industry and multi-stakeholder initiatives on AI and contributes its learnings to the development of best practice (e.g. via public or industry dialogue) • The company is equipped to respond to AI incidents controversies to different stakeholders (e.g. public, media, regulators, customers)
Performance rating	8. Public reporting on AI	• The company reports annually on its AI policies and processes • The company reports actions taken to mitigate risks and impacts • AI reporting is presented clearly in a manner accessible to non-experts • The company discloses on AI-related incidents in a timely manner	• The company describes future plans to address AI risks and opportunities in the short- and medium-term • The company provides examples of how it has identified adverse impacts associated with its AI usage and how it has mitigated these in line with principles of responsible AI usage • The company documents lessons learned from AI usage, risk management and AI-related incidents and shows how the lessons will be applied for continuous improvement

Good practice AI disclosure

Indicator 8 in our framework focuses on 'public reporting on AI'. Although investors will often gain useful intelligence from their direct interactions with portfolio companies on their approach to governing AI risks and opportunities, clear, consistent and comparable disclosure on AI is as important to aid investor understanding.

While many companies now disclose that AI is part of their strategy or innovation plans, and technology companies increasingly disclose their AI principles, very few disclose their supporting AI policies, risks or management processes. Disclosure is largely limited to brief explanations of a company's AI policies.

A core aspect of Railpen's approach to working in partnership with portfolio companies and other investors is to provide examples of good practice reporting to help inform what good might look like. We've therefore provided some excerpts from communications by companies that have published more comprehensive AI disclosures tailored to their sector.

We do not expect companies to reproduce this exactly, but we hope that these provide a useful guide for companies and their shareholders as to what useful reporting might look like.

IT

SAP

Global AI Ethics Policy

SAP's policy includes a glossary that defines all AI-related terminology, roles and responsibilities in non-technical language. It lists the company's AI principles in line with UNESCO recommendations and explains how each principle is implemented within SAP's operations. Compliance, due diligence and related policies are clearly linked.

To meet the challenges posed by the rapidly evolving AI technologies and align to an internationally recognized set of values, SAP has chosen to ground this new version of the Global AI Ethics Policy in the "Recommendation on the Ethics of Artificial Intelligence" by UNESCO. Our principles are:

1. Proportionality and do not harm
2. Safety and Security
3. Fairness and non-discrimination
4. Sustainability
5. Right to privacy and data protection
6. Human oversight and determination
7. Transparency and explainability
8. Responsibility and accountability
9. Awareness and literacy
10. Multistakeholder and adaptive governance and collaboration.

HEALTHCARE

Siemens Healthineers

Sustainability Report 2024

Siemens Healthineers' Sustainability Report describes the company's AI developments and how AI is being used in a range of products and innovations. It outlines nine principles of responsible AI use specific to the healthcare industry. It approaches managing AI risks within a human rights framing.

While the use of AI can improve access to healthcare and overcome staff shortages, it also requires a thorough analysis of potential human rights impacts. We are closely monitoring the developments pertaining to the EU regulation on AI (EU AI Act) and other related laws and regulations, and we support the EU AI Act's goals of promoting human-centric and trustworthy AI. Even before the EU AI Act was passed and implemented, some businesses within the Company started to implement the EU's Assessment List of Trustworthy AI (ALTAI) approach on a voluntary basis to ensure that human rights issues such as anti-discrimination and bias in the context of algorithm training are properly managed.

IT

Sony

Sustainability Report 2024

Responsible AI forms a chapter of Sony's 2024 Sustainability Report. The chapter covers the company's AI Ethics Guidelines (principles), organisational structure, approach to AI risks, stakeholder dialogue with business and government bodies, and collaboration with the Partnership on AI.

Sony established the Sony Group AI Ethics Committee in 2019 and reviews its use of AI and related research and development from a variety of viewpoints to ensure that activities are conducted appropriately from societal and ethical perspectives, in accordance with these Guidelines. [...]

In 2021, the AI Ethics Office was established to provide subject matter expertise on AI ethics to all Sony Group business units. In addition, Sony has established a communication system for AI utilization in products, services, and internal operations in Sony Group business units, to share information on AI ethics risks.

In March 2021, in accordance with the Sony Group AI Ethics Guidelines, Sony established an internal document stipulating requirements to be complied with in the commercialization process of electronics products and services. In July 2021, Sony started conducting AI ethics assessments in the product development life cycle, and has since assessed over 100 cases. Sony uses e-learning tools to promote an understanding of AI ethics among its employees and invites speakers from outside the company to discuss this issue at lectures and symposia.
Sony also acknowledges generative AI as an area that requires urgent attention and established internal guidelines governing use of generative AI tools at Sony Group Corporation in fiscal year 2023.

UTILITIES

Iberdrola

Policy on the Responsible Development and Use of Artificial Intelligence Tools

Iberdrola's policy describes the company's AI principles in line with OECD recommendations. It states how the policy is updated, implemented and governed by dedicated internal teams and overseen by the Audit and Risk Supervision Committee.

The Digital Transformation Division [...] shall supervise compliance with the provisions of this Policy and regularly report to the Audit and Risk Supervision Committee thereon. Similarly, the Digital Transformation Division [...] shall review this Policy at least once per year to ensure that the content thereof conforms to the ongoing progress, innovations, risks and regulatory changes that are occurring in the area."

IT

Salesforce

Artificial Intelligence Acceptable Use Policy

This policy applies to customers' (end-users) usage of Salesforce's CRM software, or any software used in combination with it. It prohibits using AI for specific uses, such as significant decision-making, individualised advice, predicting protected characteristics, and generating harmful or offensive content.

Customers may not use a Covered AI Service, nor allow their users or any third party to use a Covered AI Service, for the following:

I. Automated Decision-Making Processes with Legal Effects

   a. As part of an automated decision making process with legal or similarly significant effects, unless Customer ensures that the final decision is made by a human being. [...]

CONSUMER DISCRETIONARY

Best Buy

Policy on the Use of Generative AI Content

In the retail sector where AI use is rarely disclosed, Best Buy Canada's policy describes how, where and why the company uses AI on its website, how generated content on its website can be identified, and outlines privacy, accuracy and ethical considerations as they apply to users.

Right now, we leverage AI-generated content to help create blog posts, helpful articles that explain product category assortments, and overviews of products sold by our Marketplace seller partners.

We've designed an online experience where it will always be clear if content was generated by an AI tool. For example, if a blog post was created with the help of generative AI, the author of the article will be listed as "Best Buy (assisted with AI).

Engagement questions

If company disclosures don't align with the AIGF, investors can use the following engagement questions, across each of the framework's pillars, to better understand company practices and to encourage greater transparency and reporting.

Pillar 1: Governance

- How is AI oversight structured at the senior leadership or board level?
- Is there a senior leader and/or board member with AI expertise? If not, how is expertise strengthened through internal deep dives or external training?
- Who is responsible for managing AI deployment within the company?
- Is there a cross-functional committee overseeing AI use across departments?
- Does the company have a standalone AI policy or is it integrated into broader governance frameworks?
- How does the company ensure the policy isinformed by internal and external expertise, particularly on the topic of data management?

Pillar 2: Strategy

- How has the company assessed AI's impact on its business model?
- In what ways does AI influence strategic decision making?
- Are AI experts involved in shaping the company's strategic direction?

Pillar 3: Risk manangement

- What processes are in place to identify and assess AI-related risks across operations and supply chains?
- How does the company ensure that AI risks are regularly updated and integrated into enterprise risk management?
- Are internal and/or external audits conducted to review AI risk management?
- How does the company engage stakeholders (employees, customers, suppliers) on AI usage?
- How does the company respond to AI-related incidents?

Pillar 4: Performance reporting

- What information does the company disclose publicly about its AI policies, practices, and incidents?
- Can the company share any lessons learned from AI implementation and how these have informed risk mitigation or improvements?

Policy advocacy

AI is a system-wide issue, meaning its effective management requires constructive dialogue and collaboration with a range of stakeholders. As with other ESG issues, policymakers play a vital role in creating the right framework for corporate and market activity and behaviour on AI. However, the speed of AI adoption and its evolving capabilities means policymakers and regulators are often playing catch-up with industry developments.

Investors, given their unique insight and experiences, particularly around good governance, have a crucial role to play in helping to shape policy in this area. We've therefore summarised the major regulatory developments on AI by jurisdiction, as we recognise that this most closely aligns with how investors consider whether and how to undertake public policy activity (to the extent they do so).

Regulatory developments in AI

Globally, many countries are now developing guidance to ensure that AI developments do not infringe civil rights and democracy. Regulatory efforts highlight the tension between encouraging innovation and maintaining guardrails against the unsafe use of AI.

Significant developments for businesses include:

- **EU**: The EU Artificial Intelligence Act, which came into force in 2024, is the most comprehensive AI legislation to date, governing the development, marketing, and use of AI. The Act has differentiated obligations for AI providers (developers), deployers, importers and distributors operating within the EU. Activities are categorised by the risk they pose[120,121]:
  - **Unacceptable risk**: Systems that manipulate decisions; use specific characteristics (e.g. age, race, economic status) to exploit vulnerabilities or in biometric categorisation; social scoring; inferring emotional states in the workplace or education.
  - **High risk**: Biometrics; safety components in critical infrastructure; access to education, training, and employment (including recruitment, promotion and performance evaluation); essential services (including credit scoring and life or health insurance pricing).
  - **Limited risk**: Narrow procedural tasks; refining activities previously completed by humans; detection and decision-making which is not meant to influence without proper human review; preparatory tasks.
  - **Minimal or no risk**: Systems with no risk to safety, privacy, or rights are unregulated.

  Obligations and penalties, including significant fines, are differentiated by risk level[122]. The Act also includes specific responsibilities for providers of general purpose AI, such as publishing training data, incident tracking, and compliance with copyright. The Council of Europe also led the development of the first ever legally binding international treaty on AI, signed by member states and 10 external countries, aiming to ensure AI developments are consistent with human rights, democracy and the rule of law[123].

- **United States**: Currently, the US has no federal legislation to regulate AI, although some states such as Colorado and Illinois have legislation covering algorithmic discrimination in recruitment and employment applying to a wide range of businesses. California's AI Bill, in force from 2026, addresses transparency and safety in large-scale AI systems, although the Bill effectively pertains only to a handful of large developers[124].

- **United Kingdom**: The UK's AI strategy is focused on the potential of AI investment to boost the UK economy, provide jobs, and deliver public services more efficiently. The Government's proposed AI Opportunities Action Plan, backed by large tech companies, aims to foster innovation and increase the capacity of domestic AI developers. Investors should closely monitor the strategy and engage with consultations to promote the responsible development and use of AI[125,126].

- **China**: Several AI-related laws are in effect, covering specific use-cases such as algorithmic recommendations, GenAI, and deepfakes, as well as requiring ethical review of AI services. The regulations apply to anyone providing services to the public, including foreign businesses[127].

- **Other jurisdictions**: As listed by the International Association of Privacy Professionals' Global AI Law and Policy Tracker, countries across the globe are developing rules to govern AI use[128]. Most concern public sector use, although some specifically cover business applications. By and large, these are non-binding: they consist of policy guidelines, voluntary frameworks or bills in the drafting stage.

Policy advocacy by investors on AI is in its infancy. However, by seeking to influence policy debates, investors can help ensure regulation accounts for investor perspectives and reinforce key elements of responsible AI usage to help investors make informed decisions about their holdings.

At a minimum, investors should stay up-to-date on the evolving policy landscape and seek participation in roundtables, public consultations, and collaborative statements.

Although not AI-specific, we also recommend that investors follow the guidance provided by the ICGN in its 2023 Systemic Stewardship and Public Policy Advocacy Toolkit[h], which Railpen worked on with ICGN and members of the ICGN Global Stewardship Committee. This guidance suggests that investors should do the following:

- Assess both the need for advocacy and available resources
- Create a strategy
- Develop the public policy approach
- Implement the plan
- Track progress
- Report on the plan.

Although policy advocacy can feel uncomfortable and unfamiliar for investment practitioners, we encourage investors to leverage the experience, expertise and relationships of their corporate affairs team and/or support through a specialist third-party supplier.

The investor perspective is of particular value in relation to the following:

- **Transparency** – increasing transparency and disclosure on approaches to AI through reporting. Ensuring end-users know when and how AI is being used, and that shareholders get the kind of information they need in both written and informal communications to help them understand a company's approach to and governance of AI.
- **Skills and education** – upskilling of workers in AI. Ensuring decision-makers and end-users are empowered to make decisions on AI use and have a working knowledge of its capabilities and risks. This includes training that is not 'one and done' but is instead refreshed and refined over time.
- **Principles and processes** – promoting safe and ethical standards of AI use as outlined by our summarised responsible AI principles (see page 29). Ensuring these principles are put into practice effectively, consistently and transparently.
- **Governance** – creating robust accountability, governance and management structures for AI usage. This could include incentivising effective board composition and structure, considering whether there are issues preventing the implementation of robust systems, processes and controls at companies on AI.
- **Risk identification** – defining high-risk uses of AI across business sectors and their differentiated responsibilities. This may mean formal prohibition of what are deemed to be unacceptable risks, or requesting additional reporting on oversight processes or responses to AI crises.
- **Risk management** – mandating regular and robust processes of risk management.

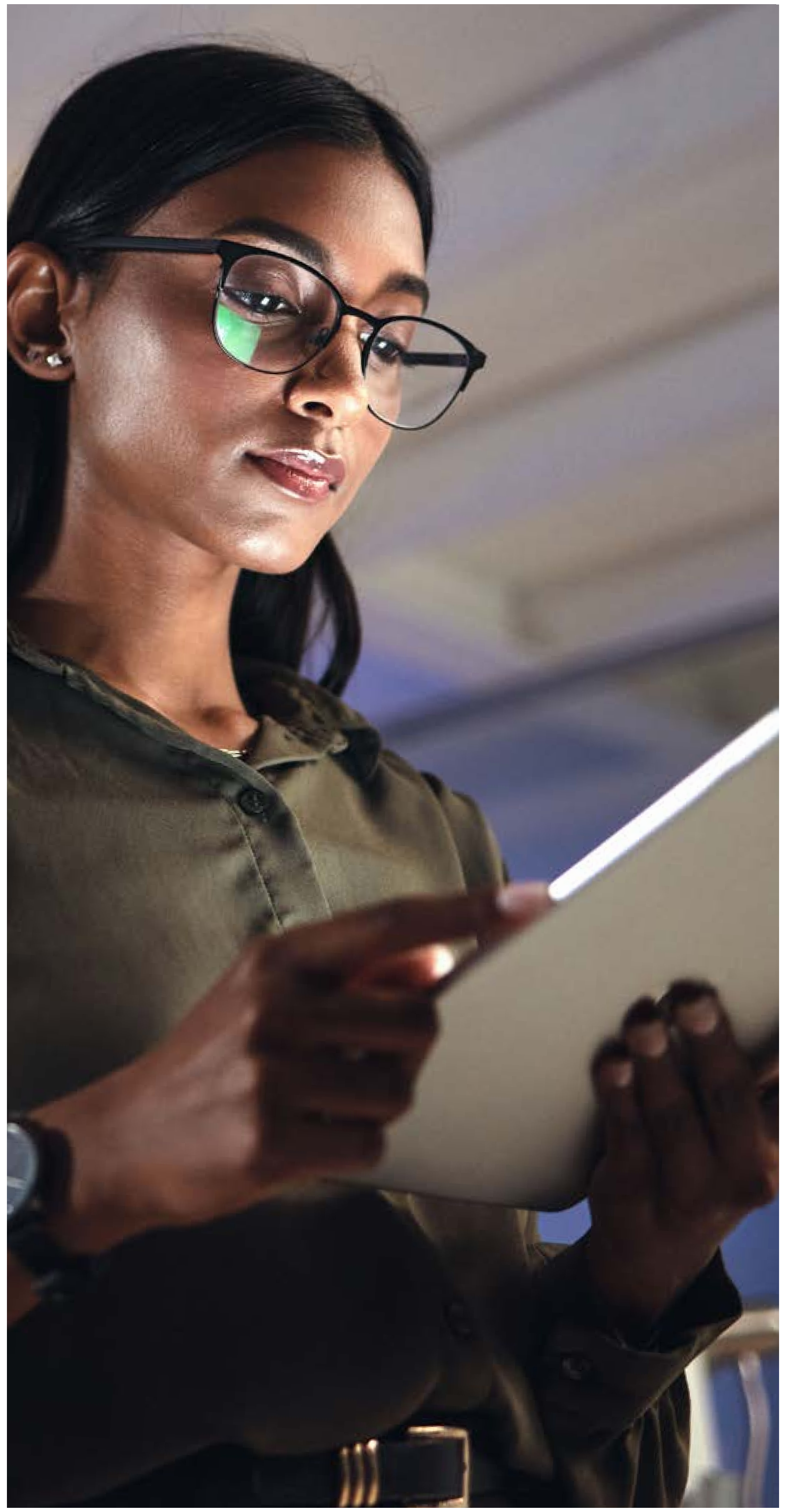
While many investors will have their own public policy strategies for AI and other issues, collaborating with others', including policymakers, can offer significant additional benefits. In our experience, policymakers often appreciate it when an industry or group presents a unified, clear and consistent voice. Therefore, we encourage investors to consider joining membership associations or existing collective initiatives that are actively involved in policy advocacy.

Good practice case study

The work of the Investor Alliance for Human Rights (IAHR), through its Tech and AI Policy workstream, provides some good examples of a collaborative approach to policy advocacy on AI issues. For instance:

- IAHR filed a public submission with the U.S. National Telecommunications and Information Administration (NTIA) to highlight the imperative for AI policy to include mandatory human rights due diligence (HRDD) requirements, prohibitions and restrictions on AI use[129].
- 149 institutional investors signed the Alliance's Investor Statement in Support of Digital Rights Regulations which provided investor recommendations to the EU AI Act[130].

[h] Systemic Stewardship & Public Policy Advocacy Toolkit



CONCLUSIONS & RECOMMENDATIONS

The rapid adoption of AI technologies in business environments brings with it considerable risks and opportunities.

AI's positive and transformative power spans various sectors; however the associated risks must not be overlooked. The rising incidents and controversies around AI, along with their financial repercussions, underscore the need for robust governance frameworks to manage these risks effectively.

Currently, there's a clear gap between the recognition of AI risks and the preparedness of companies to manage them. Despite the high anticipation of threats associated with GenAI, most companies lack the necessary governance structures to mitigate the risks. Regulation is behind the rapid pace of AI development, with current policies focusing more on growth and rather than safety.

Investors are pivotal in bridging this gap by encouraging responsible AI use and effective governance practices. The AIGF in this report aims to provide companies and investors with the necessary tools to address current AI risks, while also preparing for future uncertainties and opportunities. By cultivating robust governance practices, organisations can enhance their resilience in the face of technological, regulatory, and societal changes.
In summary, this report encourages investors to support collective action on AI through the following:

1. **Conduct a high-level risk assessment** to determine portfolio risks from AI using the different criteria presented in this report.
2. **Engage with priority companies using our AIGF** as a basis for assessing how companies are managing AI risks, and our guidance on good practice disclosure and potential engagement questions to support meaningful dialogue.
3. **Consider engaging in policy advocacy** around the responsible use of AI to close the gap between regulation and its rapid evolution, and ensure the unique and useful investor perspective is heard by policymakers.

The responsible development and use of AI in a way that benefits processes, people and profits requires collective efforts from investors to work in partnership with companies and policymakers. Only then can we harness the significant opportunities AI can offer while mitigating its most significant risks .

REFERENCES

1 McKinsey (2024). The state of AI in early 2024 | McKinsey.

2 Deloitte and USC Marshall (2024). Largest Companies View AI as a Risk Multiplier.

3 McKinsey (2024). The state of AI in early 2024 | McKinsey.

4 Deloitte and USC Marshall (2024). Largest Companies View AI as a Risk Multiplier.

5 Stanford Institute for Human-Centered Artificial Intelligence (HAI). Artificial Intelligence Index Report (2024). hai_ai-index-report-2024-smaller2.pdf.

6 Deloitte (2024). Providing insurance coverage for artificial intelligence may be a blue ocean opportunity. Risk insurance for AI coverage | Deloitte Insights.

7 Center for International Governance Innovation (2023). AI-Related Risk, The Merits of an ESG-Based Approach to Oversight no.279.pdf.

8 Reuters (2024). Italy fines OpenAI over ChatGPT privacy rules breach | Reuters.

9 Riskonnect (2023). 2023 New Generation of Risk Survey. The New Generation of Risk: Are You Ahead of the Curve or Behind the Pack? · Riskonnect.

10 WEF (2024). WEF_Responsible_AI_Playbook_for_Investors_2024.pdf.

11 Information Commissioner's Office (n.d.). Part 1 The basics of explaining AI: Definitions. Definitions | ICO.

12 OECD (n.d.). OECD AI Principles Overview. AI Principles Overview - OECD.AI.

13 Escott (2017). What are the 3 types of AI? A guide to narrow, general, and super artificial intelligence | Codebots.

14 Roger (2021). What are the branches of Artificial Intelligence? | H2K Infosys Blog.

15 IBM (2023). AI vs. machine learning vs. deep learning vs. neural networks: What's the difference? AI vs. Machine Learning vs. Deep Learning vs. Neural Networks | IBM.

16 Digital Adoption (2024). The Most Important Branches of Artificial Intelligence You Need To Know

17 McKinsey (2018). Notes from the AI frontier: Modeling the impact of AI on the world economy. Modeling the global economic impact of AI | McKinsey.

18 Deloitte and USC Marshall (2024). Largest Companies View AI as a Risk Multiplier.

19 UN Global Compact Network Germany (2024). Artificial Intelligence and Human Rights: Recommendations for companies. Full Report_AI and Human Rights_Recommendations for Companies.pdf.

20 McKinsey (2018). Notes from the AI frontier: Modeling the impact of AI on the world economy. mgi-notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy-september-2018.ashx.

21 Ormsby (2019). Uber fatality unveils AI accountability issues. Uber fatality unveils AI accountability issues - Lawyers Weekly.

22 Yagoda (2024). Airline held liable for its chatbot giving passenger bad advice – what this means for travellers. https://www.bbc.com/travel/article/20240222-air-canada-chatbot-misinformation-what-travellers-should-know.

23 Vincent (2024). How much electricity does AI consume? How much electricity do AI generators consume? – The Verge.

24 Goldman Sachs (2024). AI is poised to drive 160% increase in data center power demand. AI is poised to drive 160% increase in data center power demand | Goldman Sachs.

25 Gordon (2024). AI Is Accelerating the Loss of Our Scarcest Natural Resource: Water. https://www.forbes.com/sites/cindygordon/2024/02/25/ai-is-accelerating-the-loss-of-our-scarcest-natural-resource-water/.

26 Nguyen (2024). AI is making Google and Microsoft big contributors to climate change. https://qz.com/ai-google-microsoft-climate-change-data-center-energy-1851589453.

27 Ghaffary (2024). Big Tech's Climate Goals At Risk From Massive AI Energy Demands. Big Tech's Climate Goals At Risk From Massive AI Energy Demands - Bloomberg.

28 McGovern and Brandford (2023). The Cloud vs. drought: Water hog data centers threaten Latin America, critics say.

29 Rogoway (2023). The Dalles settles public records lawsuit over Google's data centers, will disclose water use to The Oregonian/OregonLive - oregonlive.com.

30 UN OHCHR (2024). Racism and AI: "Bias from the past leads to bias in the future". Racism and AI: "Bias from the past leads to bias in the future" | OHCHR.

31 Dastin (2018). Insight - Amazon scraps secret AI recruiting tool that showed bias against women. Insight - Amazon scraps secret AI recruiting tool that showed bias against women | Reuters.

32 BBC News (2018). Amazon scrapped 'sexist AI' tool. Amazon scrapped 'sexist AI' tool - BBC News.

33 Wiggers (2020). Researchers find evidence of racial, gender, and socioeconomic bias in chest X-ray classifiers. Researchers find evidence of racial, gender, and socioeconomic bias in chest X-ray classifiers | VentureBeat.

34 AIAAIC (2021). US mortgage approval algorithm more likely to reject people of colour. AIAAIC - US mortgage approval algorithm discrimination.

35 Thoropass (n.d.). AI data breach: Understanding their impact and protecting your data. AI data breach: Understanding their impact and protecting your data - Thoropass.

36 White (2024). Real-Life Examples of How AI Was Used to Breach Businesses. Real-Life Examples of How AI Was Used to Breach Businesses - New.

37 AI Incident Database (2023). Incident 443: ChatGPT Abused to Develop Malicious Softwares. Incident 443: ChatGPT Abused to Develop Malicious Softwares.

38 Anand (2024). The High Cost of Non-Compliance: Penalties Issued for AI under Existing Laws.

39 Lyons (2020). Clearview AI's client list includes 2,200 organizations spanning law enforcement to universities. Clearview AI's clients include 2,200 organizations from law enforcement to universities - The Verge.

40 Gesser et al. (2023). Legal Risks of Using AI Voice Analytics for Customer Service. Legal Risks of Using AI Voice Analytics for Customer Service – Debevoise Data Blog.

41 ISS Insights (2024). AI and Safety: Mind the Accumulation of Mass-Market Risks.

42 Bernier (2023). Cruise recalls 950 self-driving vehicles over software glitch. Cruise recalls 950 self-driving vehicles over software glitch – Houston Public Media.

43 Tri-College Libraries (n.d.). Generative AI in Higher Education (BMC). Hallucinations and Deep Fakes - Generative AI in Higher Education (BMC) - Research Guides at Tri-College Libraries.

44 Yeo (2024). Viral fashion company Selkie is being slammed for using AI art. Viral fashion company Selkie is being slammed for using AI art | Mashable.

45 Fear (2024). Lego is the latest brand to apologise for using AI art. Lego is the latest brand to apologise for using AI art | Creative Bloq.

46 Shaffi (2023). Bloomsbury admits using AI-generated artwork for Sarah J Maas novel. Bloomsbury admits using AI-generated artwork for Sarah J Maas novel | Books | The Guardian.

47 Weatherbed (2024). Artists are making creative companies apologize for using AI. Artists keep rallying against creative companies for using AI - The Verge.

48 Vassen (2024). Calgary Farmers' Market calls negative reaction to AI art a 'tempest in a teapot. Calgary Farmers' Market calls negative reaction to AI art a 'tempest in a teapot' - Calgary | Globalnews.ca.

49 AIAAIC (2024). "Dangerous" AI-generated mushrooms flood Google. AIAAIC - "Dangerous" AI-generated mushrooms flood Google.

50 Fairly AI (2023). Fairly AI | Fairly AI Submission to the United Nations' Call for Papers: Data Pollution in the Age of Generative AI.

51 Lapienytė (2024). AI tools can lead to severe mushroom poisoning. AI tools can lead to severe mushroom poisoning | Cybernews.

52 Damiani (2019). A Voice Deepfake Was Used To Scam A CEO Out Of $243,000.

53 Chen (2024). Finance worker pays out $25 million after video call with deepfake 'chief financial officer'. Finance worker pays out $25 million after video call with deepfake 'chief financial officer' | CNN.

54 The Guardian (2016). Microsoft 'deeply sorry' for racist and sexist tweets by AI chatbot | Artificial intelligence (AI) | The Guardian.

55 Schwartz (2019). In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation. In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation - IEEE Spectrum.

56 El Atillah (2023). Man ends his life after an AI chatbot 'encouraged' him to sacrifice himself to stop climate change. Man ends his life after an AI chatbot 'encouraged' him to sacrifice himself to stop climate change | Euronews.

57 Hoffman (2024). Florida mother files lawsuit against AI company over teen son's death: "Addictive and manipulative". Florida mother files lawsuit against AI company over teen son's death: "Addictive and manipulative" – CBS News.

58 Milmo (2025). Apple suspends AI-generated news alert service after BBC complaint. Apple suspends AI-generated news alert service after BBC complaint | Apple | The Guardian.

59 Milmo (2023). Two US lawyers fined for submitting fake court citations from ChatGPT. Two US lawyers fined for submitting fake court citations from ChatGPT | ChatGPT | The Guardian.

60 Bitlaw (n.d.) AI Hallucinations (Why would I lie?). AI Hallucinations (Why would I lie?) (BitLaw).

61 AI Incident Database (2023). Incident 545: Chatbot Tessa gives unauthorized diet advice to users seeking help for eating disorders.

62 AIAAIC (n.d). nH Predict post-acute care predictions. AIAAIC - NaviHealth nH Predict post-acute care predictions.

63 Talia (2024). Algorithms That Can Deny Care, and a Call for AI Explainability. https://www.computer.org/csdl/magazine/co/2024/07/10574502/1Y7B8DpFlw4.

64 Engel et al. (2024). Code is law: how COMPAS affects the way the judiciary handles the risk of recidivism. https://link.springer.com/article/10.1007/s10506-024-09389-8.

65 Los Angeles Daily News (2024). 'Blade Runner 2049' producers sue Elon Musk and Tesla over AI image at robotaxi event. 'Blade Runner 2049' producers sue Elon Musk, Tesla.

66 Authors Guild (2023). The Authors Guild, John Grisham, Jodi Picoult, David Baldacci, George R.R. Martin, and 13 Other Authors File Class-Action Suit Against OpenAI. https://authorsguild.org/news/ag-and-authors-file-class-action-suit-against-openai/.

67 Mishcon de Reya (2024). Generative AI – Intellectual property cases and policy tracker. Generative AI – IP cases and policy tracker | Mishcon de Reya.

68 Tencer (2024). As Suno and Udio admit training AI with unlicensed music, record industry says: 'There's nothing fair about stealing an artist's life's work.' As Suno and Udio admit training AI with unlicensed music, record industry says: 'There's nothing fair about stealing an artist's life's work.' – Music Business Worldwide.

69 IPPR (2024). Up to 8 million UK jobs at risk from AI unless government acts, finds IPPR | IPPR.

70 Ibid.

71 De Stefano and Taes (2022). Algorithmic management and collective bargaining. https://journals.sagepub.com/doi/full/10.1177/10242589221141055.

72 Global Partnership on Artificial Intelligence (2024). Fairwork Amazon Report 2024. Fairwork-Amazon-Report_June24-2.pdf.

73 University of Oxford (2024). The hidden cost of AI: In conversation with Professor Mark Graham. The hidden cost of AI: In conversation with Professor Mark Graham | University of Oxford.

74 Al-Sibai (2024). The Tesla Workers Dealing With Autopilot Data Are Treated So Strangely You Won't Believe It. https://futurism.com/the-byte/tesla-autopilot-workers-treatment.

75 Williams et al. (2022). The Exploited Labor Behind Artificial Intelligence.

76 Doellgast et al. (2023). AI in contact centers: Artificial intelligence and algorithmic management in frontline service workplaces. AI in Contact Centers.

77 Workday (2024). Global Study: Closing the AI trust gap. 2024 Global Study: Closing the AI trust gap.

78 AIAAC (2024). Australian voice artists lose work to AI clones. AIAAIC - Australian voice artists lose work to AI clones.

79 Horton (2023). 'Embrace it or risk obsolescence': how will AI jobs affect Hollywood? 'Embrace it or risk obsolescence': how will AI jobs affect Hollywood? | Film industry | The Guardian.

80 Amidi (2024). New Report Confirms Worst Fears: AI Will Disrupt Countless Animation Jobs Over Next 3 Years.

81 Greenhouse (2024). 'Constantly monitored': the pushback against AI surveillance at work. 'Constantly monitored': the pushback against AI surveillance at work | Artificial intelligence (AI) | The Guardian.

82 Personnel Today (2022). Estée Lauder staff win payout after being 'sacked by algorithm'. Sacked by algorithm: Estée Lauder staff win out-of-court payout.

83 Faragher (2023). Uber fined after 'robo-firing' drivers. https://www.personneltoday.com/hr/uber-robo-firing-fine/.

84 Derry et al. (2024). Does the Use of AI in the Hiring Process Expand Who Can Be Sued for Discrimination?: One Federal Court in California Says Yes. https://www.paulhastings.com/insights/client-alerts/does-the-use-of-ai-in-the-hiring-process-expand-who-can-be-sued-for.

85 McKinsey (2024). The State of AI in early 2024. The state of AI in early 2024 | McKinsey.

86 Center for American Progress (2024). To Implement AI Responsibly, Third-Party Deployments Must Require Safeguards. To Implement AI Responsibly, Third-Party Deployments Must Require Safeguards - Center for American Progress.

87 Business Software Alliance (2023). AI Developers and Deployers: An Important Distinction.

88 Osborne Clarke (2024). EU AI Act's 'deployers' definition has wide-ranging significance for life sciences. EU AI Act's 'deployers' definition has wide-ranging significance for life sciences - Osborne Clarke | Osborne Clarke.

89 AIprophet (2024). How AI Is Transforming the S&P 500's Eleven Sectors. How AI Is Transforming the S&P 500's Eleven Sectors | LinkedIn.

90 Precedence Research (2024). Artificial Intelligence in Healthcare Market Size, Share and Trends 2024 to 2034. Artificial Intelligence in Healthcare Market Size, Report 2034.

91 EY Global (2022). Why AI will redefine the financial services industry in two years. https://www.ey.com/en_gl/insights/innovation/why-ai-will-redefine-the-financial-services-industry-in-two-years.

92 Google Cloud (2024). Google Cloud Shares New Research on 2024 Outlook on Generative AI in Retail. Google Cloud Shares New Research on 2024 Outlook on Generative AI in Retail - Jan 11, 2024.

93 Penrod (2023). A third of utilities have begun to pilot generative AI for customer service, other uses: report. A third of utilities have begun to pilot generative AI for customer service, other uses: report | Utility Dive.

94 IBM (2024). New IBM Study Data Reveals 74% of Energy & Utility Companies Surveyed Embracing AI. New IBM Study Data Reveals 74% of Energy & Utility Companies Surveyed Embracing AI.

95 Nvidia (2024). State of AI in Telecommunications: 2024 Trends Survey Report. State of AI in Telecommunications: 2024 Trends.

96 Ground (2024). The Rise of AI and ESG. The Rise of AI and ESG | Capital Group.

97 Chow et al. (2019). Investors' Expectations on Responsible Artificial Intelligence and Data Governance. investors-expectations-on-responsible-artificial-intelligence-and-data-governance.pdf.

98 LGIM (2023). CIO autumn update: An AI inflection Point. CIO Autumn Update.

99 LGIM (2023). Q4 2023 Quarterly engagement report.

100 LGIM (2024). LGIM's voting intentions for 2024. LGIM Blog: LGIM's voting intentions for 2024.

101 Riddick (2024). Responsible AI – What Role Can an Investor Play?

102 Brunel Pension Partnership (n.d.). GSK, Cyber & AI, responsible use of AI policy. GSK, Cyber & AI, responsible use of AI policy - Brunel Pension Partnership.

103 Cave et al. (2024) Unveiling Key Trends in AI Shareholder Proposals. Unveiling Key Trends in AI Shareholder Proposals - FTI Strategic Communications.

104 Hudson (2024). Investors Eye Proxy Efforts on Climate, AI, Diversity in 2025.

105 Gambetta (2025). AI resolutions: Investors target companies on net-zero goals and water. AI resolutions: Investors target companies on net-zero goals and water.

106 Maiden (2024). AI ethics proposal attracts support among Netflix shareholders. https://www.governance-intelligence.com/shareholders-activism/ai-ethics-proposal-attracts-support-among-netflix-shareholders.

107 Cave et al. (2024) Unveiling Key Trends in AI Shareholder Proposals. Unveiling Key Trends in AI Shareholder Proposals – FTI Strategic Communications.

108 World Economic Forum (2024). Responsible AI Playbook for Investors. WEF_Responsible_AI_Playbook_for_Investors_2024.pdf.

109 World Benchmarking Alliance (2023). Augmenting Ethical AI: 2023 Progress Report on the Collective Impact Coalition for Digital Inclusion. Digital-CIC-2023-Progress-Report_.pdf.

110 CSIRO (2024). The intersection of Responsible AI and ESG: A Framework for Investors. Responsible AI ESG Framework for investors - CSIRO.

111 UNESCO (2021). Recommendation on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000380455.

112 OECD (n.d.). OECD AI Principles Overview. AI Principles Overview - OECD.AI.

113 Partnership on AI (n.d.). Our Resource Library. Resource Library - Partnership on AI.

114 Responsible Artificial Intelligence Institute (n.d.). Accelerating Responsible AI Adoption. Home - Responsible AI.

115 CFA Institute (2022). Ethics and Artificial Intelligence in Investment Management: A Framework For Professionals. Ethics-and-Artificial-Intelligence-in-Investment-Management_Online.pdf.

116 ICGN (2024). Artificial intelligence: An engagement guide. ICGN Investor Viewpoint - Artificial Intelligence - An engagement guide (2).pdf.

117 Responsible Artificial Intelligence Institute (2024). AI's Impact on Our Sustainable Future: A Guiding Framework for Responsible AI Integration Into ESG Paradigms. AIs Impact on Our Sustainable Future White Paper V1.pdf.

118 Responsible Artificial Intelligence Institute (n.d.). Frequently asked questions. FAQs - Responsible AI.

119 World Benchmarking Alliance (2023). Augmenting Ethical AI: 2023 Progress Report on the Collective Impact Coalition for Digital Inclusion. Digital-CIC-2023-Progress-Report_.pdf.

120 Forvis Mazars (n.d.). EU AI Act: different risk levels of AI systems. EU AI Act: different risk levels of AI systems – Forvis Mazars – Ireland.

121 EU Artificial Intelligence Act (2024). High-level summary of the AI Act. High-level summary of the AI Act | EU Artificial Intelligence Act.

122 Levi et al. (2024). The EU AI Act: What Businesses Need To Know. The EU AI Act: What Businesses Need To Know | Insights | Skadden, Arps, Slate, Meagher & Flom LLP.

123 Council of Europe (n.d.). The Framework Convention on Artificial Intelligence. The Framework Convention on Artificial Intelligence – Artificial Intelligence.

124 Davies (2024). A big win for the EU? How California's new AI bill compares to the EU AI Act. A big win for the EU? How California's new AI bill compares to the EU AI Act | Euronews.

125 Department for Science, Innovation & Technology (2025). AI Opportunities Action Plan. AI Opportunities Action Plan – GOV.UK.

126 McMahon et al. (2025). PM plans to 'unleash AI' across UK to boost growth. Artificial Intelligence: Plan to 'unleash AI' across UK revealed - BBC News.

127 Latham & Watkins (2023). China's New AI Regulations.

128 IAPP (2024). Global AI Law and Policy Tracker. global_ai_law_policy_tracker.pdf.

129 Regulations.gov (2023). Comment on FR Doc # 2023-07776. Regulations.gov.

130 Investor Alliance for Human Righjts (2023). Investor Statement in Support Of Digital Rights Regulations European Union Artificial Intelligence Act. FINAL Investor Statement AI Act w-signatories 2-14-23_0.pdf.

Appendix

Table 1: **AI classification by capability**

Capability	Definition	Emphasis in this report
Artificial narrow intelligence or 'weak' AI	Systems that autonomously perform a specific predefined task when prompted using human-like capabilities.	High
Artificial general intelligence or 'strong' AI	Systems that mimic human perception and understanding, independently learn new tasks, and form competencies and connections across different domains and functions of AI.	Low
Artificial super-intelligence or 'super' AI	Systems which supersede human performance, due to enhanced memory, speed and processing capabilities.	Low

Table 2: **AI classification by functionality**

Functionality	Definition	Emphasis in this report
Reactive AI	Systems which respond autonomously to a limited set of inputs. Reactive AI does not have memory-based functionality, i.e. it cannot 'learn' from previous experience.	Medium
Limited memory	Systems which use historical data and previous experiences to improve their capability, allowing them to increase in accuracy based on 'learning experience'.	High
Theory of mind	A theoretical category of systems that can understand human thought processes (beliefs, emotions, needs).	Low
Self-aware	A hypothetical category describing a system evolved to be so akin to the human brain that it develops self-awareness.	Low

List of abbreviations	
AI	Artificial Intelligence
AIGF	AI Governance Framework
DL	Deep Learning
GDPR	General Data Protection Regulation
GenAI	Generative Artificial Intelligence
GPU	Graphics Processing Unit
IAHR	Investor Alliance for Human Rights
IP	Intellectual Property
LLM	Large Language Model
ML	Machine Learning
NLP	Natural Language Processing
ESG	Environmental, Social, and Governance
S&P	Standard & Poor's
OECD	Organisation for Economic Co-operation and Development
UNESCO	United Nations Educational, Scientific and Cultural Organisation
ICGN	International Corporate Governance Network

Disclaimer

The information contained in this document is provided solely for informational purposes and has been produced by Railpen Limited working in collaboration with Chronos Sustainability. While every effort has been made to ensure the accuracy and completeness of the information, no representations, or warranties (whether express or implied) are made about the completeness, accuracy, reliability, suitability, or availability of the information contained herein. This document does not constitute legal, financial, or professional advice and should not be relied upon as such. Railpen Limited and Chronos Sustainability disclaim any and all liability for any loss or damage, whether direct, indirect, consequential, or otherwise, arising from the use or reliance on the information contained herein.

The views expressed within this document are those of Railpen Limited and Chronos Sustainability at the date of publication, which are subject to change. The information contained within this document does not constitute investment advice and should not be treated as such.

Railpen Limited is a wholly owned subsidiary of Railways Pension Trustee Company Limited, registered in England and Wales under company number 02315380 whose registered office is at 7 Devonshire Square, London, EC2M 4YH

✉ 7 Devonshire Square, Lonodn, EC2M 4YH
@ SO@railpen.com

RAILPEN

CHRONOS
INTELLIGENT SUSTAINABILITY