In collaboration with
CPP Investments Insights Institute

WORLD
ECONOMIC
FORUM

Responsible AI
Playbook for Investors

WHITE PAPER

JUNE 2024

Contents

Foreword

Cathy Li
Head, AI, Data and
Metaverse; Member of
the Executive Committee,
World Economic Forum

Judy Wade
Managing Director and Head,
Strategy Execution and
Relationship Management,
CPP Investments

When the World Economic Forum and CPP Investments Insights Institute embarked on this partnership in mid-2022, the broad relevance of responsible artificial intelligence (RAI) to investors was still in question. The role of investors in accelerating the adoption of RAI, particularly given other expectations placed on companies in the areas of climate change, equity, diversity and inclusion, and cybersecurity, was also unclear.

With the launch of ChatGPT in late 2022, both the opportunities and risks of AI came into sharper focus. In short order, the importance of adopting RAI principles and policies was no longer in doubt.

What, then, is the specific function of investors in advancing the adoption of those principles and policies?

As articulated by Amara's Law, we often overestimate the effect of a technology in the short term and underestimate the effect in the long term. This adage reminds us that the early excitement surrounding generative AI's potential, while justified, ought to be balanced with a longer-term, prudent approach to foundational concerns. The broader business community is already struggling with risks such as data breaches, privacy loss, job loss, ethical challenges, misinformation and disinformation.

To protect their portfolios' stability both now and in the future, investors should tackle both these immediate issues as well as the deeper implications of AI. This starts with establishing strong governance frameworks and clear principles and practices to integrate RAI standards into all applications.

If successful, the term "RAI" will become obsolete as high-quality, trustworthy and safe AI becomes the norm. Just as we don't distinguish between "bridges" and "bridges that don't collapse", the qualifier "responsible" will become an unspoken expectation. Today, we are at the beginning of this era, as new laws and regulations emerge to ensure that all AI applications are responsible.

This brings us back to our original question, for which there is an unequivocal answer: large investors can and should exercise the influence afforded by their capital to promote the use of RAI in their portfolios, in their work with investment partners, and in the ecosystem at large.

This white paper offers a playbook for how investors can accelerate the adoption of RAI to help drive value. While it is not a comprehensive view of all levers available, we believe it provides useful guidance, examples and concrete steps for engaging with other stakeholders.

Executive summary

Responsible artificial intelligence positions companies for success in the Fourth Industrial Revolution.

Artificial intelligence (AI) is taking a central role in the Fourth Industrial Revolution, a period defined by the fusion of technologies and a "blurring of the lines" between physical, digital and biological spheres.[1] AI is rapidly finding its way into a broad array of business-to-business and business-to-consumer applications. For large investors, ensuring all AI applications are responsible (i.e. honest, helpful and harmless) is not merely a technological upgrade but a strategic imperative. Ensuring it is developed and deployed in a manner consistent with responsible AI (RAI) principles is an important step for enhancing risk-adjusted returns and positioning businesses for success.

RAI reduces risk and promotes growth in the following ways:

– **It addresses non-regulatory business risks**. The adoption of AI carries risks of unintended consequences across ethical (social) and technological dimensions, such as new types of cyberattacks, unwanted biases, job disruptions and/or displacement, and data leaks or poisoning. A comprehensive RAI strategy can help identify and mitigate these risks.

– **It mitigates legal and regulatory risks**. A proactive RAI framework can anticipate and adapt to legal and regulatory changes, safeguarding businesses and their investors from fines and reputational harm.

– **It can improve top- and bottom-line growth**. Studies indicate RAI can increase customer trust and, therefore, engagement and retention. It can also protect brand safety, broaden revenue streams, offer procurement advantages in competitive bidding processes, and increase pricing power in the marketplace, outperforming AI systems and businesses less aligned with RAI.

To capitalize on AI's potential, investors should engage with stakeholders, including:

– **Corporate boards**: By engaging with boards of portfolio companies, direct investors can help establish (or hold directors accountable for establishing) RAI principles, policies and accountability measures.

– **Investment partners**: Asset owners can encourage investment partners to adopt AI governance in their own operations and extend it into their holdings.

– **The broader ecosystem**: Over a longer period, investors' efforts can help create an ecosystem where the benefits of RAI are well understood and adoption is ubiquitous.

A few key areas requiring further development are:

– **AI governance and standards**: A dynamic framework for AI governance is essential, as are universally accepted measures of its effectiveness. Collaborative efforts among private sector actors, academia and government (among others) can help speed up the development of these tools.

– **Education and capacity building**: To keep pace with AI advancements, stakeholders must invest in continuous learning. This includes executive education, forums for investor dialogue and public awareness initiatives.

– **Balancing speed of deployment with safety concerns**: Companies will face the temptation to develop and deploy AI rapidly in pursuit of short-term gains, treating RAI as an afterthought rather than a forethought. Without appropriate attention to RAI, hasty deployment will put long-term value at risk.

Introduction

Responsible artificial intelligence can preserve and create value for investors and other stakeholders.

> **All stakeholders for whom AI is a material risk or opportunity have an incentive to accelerate RAI's adoption. Yet this often lags behind other strategic AI priorities.**

While the rapid development of generative artificial intelligence (AI) has recently captured headlines and imaginations, the power of AI has been growing for more than 70 years.[2] The term AI was first proposed at the 1956 Dartmouth Summer Research Project on Artificial Intelligence through the introduction of the Logic Theorist – a program designed to mimic the problem-solving skills of a human.[3]

Since then, AI has been integrated into operations in a variety of industries, from healthcare to finance, manufacturing to transport. Today, the technology facilitates drug discovery, enables fraud detection, powers predictive maintenance and drives vehicles. Its influence is only expected to grow. According to a 2023 McKinsey & Company report, AI's potential impact on the global economy could be upwards of $25.6 trillion annually, which, even if inaccurate given the challenges of forecasting (particularly in fast-changing environments), indicates its potential scale.[4,5]

BOX 1 | **RAI dimensions and principles**

RAI is the practice of designing, building, deploying, operationalizing and monitoring AI systems in a manner that empowers people and businesses and impacts customers and society equitably.[6] This paper uses the most inclusive definition of RAI to cover both the ethical (social) and technical dimensions of AI, adhering to principles of validity and reliability, safety, fairness, security and resilience, accountability and transparency, explainability and interpretability, and privacy.[7]

The applicability of these principles is context dependent. For example, when AI is used in healthcare, reliability and safety might be of utmost importance, whereas when it is used in hiring processes, fairness could be the key concern. The challenge lies in finding the right balance between these principles within a given context, which often involves making conscious trade-offs.

RAI has the potential to mitigate risk and increase value for stakeholders. All stakeholders for whom AI is a material risk or opportunity (e.g. corporations developing and adopting AI, investors financing those companies and players in the broader ecosystem) have an incentive to accelerate RAI's adoption. Yet this often lags behind other strategic AI priorities.[8]

This playbook draws on research and insights from stakeholders across the AI ecosystem. It aims to help investors understand and assess the integration of RAI in their portfolios and opportunities for engagement. Investors can consider adopting elements of this playbook as they develop their own AI strategies.

1 The business case for RAI and the role of stakeholders

Investors have an incentive and an opportunity to accelerate RAI.

1.1 AI will drive both risks and profits

As AI adoption proliferates, it will become a driver of both risk (i.e. the potential for harm) and value. This means it will also become an important part of the due diligence investors perform to assess potential investments. Ultimately, how it is governed and managed will become a core oversight priority for corporate boards.

According to McKinsey & Company, organizations need to manage nine categories of risk, as identified in Table 1.

TABLE 1 — Some categories of AI risks (illustrative, non-exhaustive)

Risk category	Sub-risks	Description
Ethical and social	Privacy violations	AI systems might compromise privacy by the unauthorized use/disclosure of personal or sensitive information.
	Bias and discrimination	AI may perpetuate biases from its training data, resulting in discrimination and unfair outcomes.
	Workforce and environmental	AI deployment may amplify existing inequalities or create negative impacts on employment, job quality and the environment.
Technical and operational	Inaccurate output	AI-generated output may be false or misleading, undermining decision-making and causing adverse effects.
	Lack of explainability and accountability	AI-generated output may be hard to interpret, and it may be difficult to allocate accountability for different harms caused by AI.
	Third-party risk	Use of AI or services provided by external vendors without appropriate guardrails and/or inadvertent intellectual property (IP) leakage into the public domain can lead to negative consequences.
Security and legal	IP infringement	AI systems lead to ethical challenges when used in ways that violate copyright, patents, or other IP rights.
	Security threats	Hidden weaknesses within AI systems may create exploitable vulnerabilities, posing cybersecurity risks.
	Incorrect/ malicious use	Misusing AI for unintended purposes can lead to unintentional harm, particularly when users misunderstand its limitations. In some cases, AI may be weaponized for malicious intent, such as launching cyberattacks.

Overall, AI risks can lead to significant unintended or maliciously intended negative consequences or harms for organizations and their investors.

These harms can impact financial performance, non-financial performance, legal and compliance issues, and reputational integrity.[9]
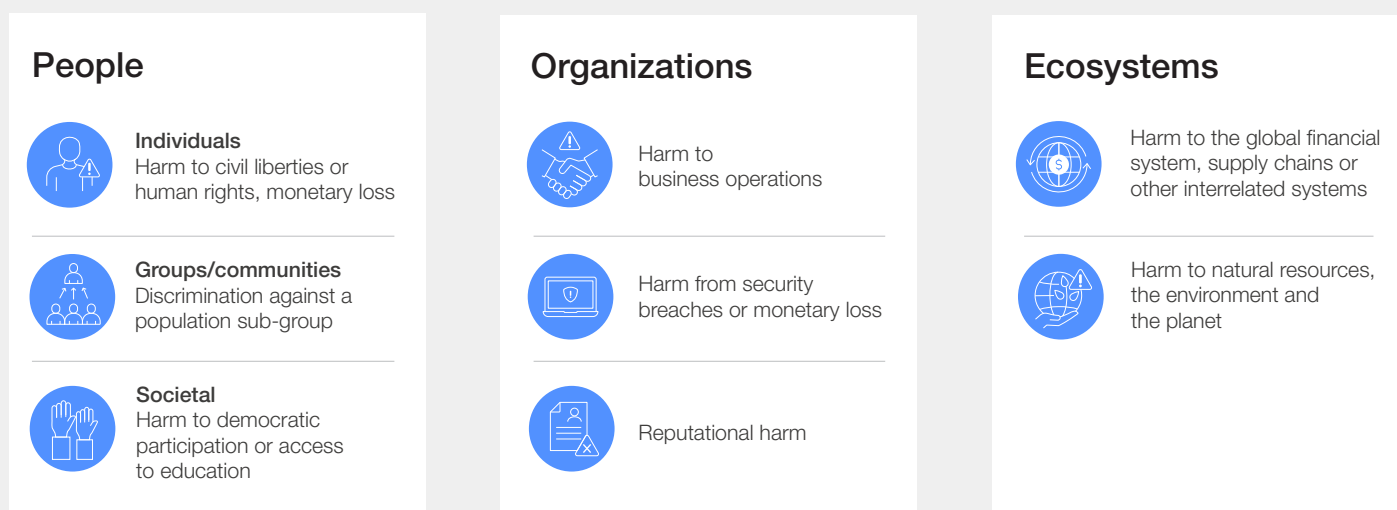
Real-world corporate setbacks underscore AI's potential to harm corporations and their investors. For instance, Zillow lost approximately $881 million on its Zillow Offers home-flipping business in 2021 when its models did not adjust in tandem with a cooling housing market.[10] The New York Times sued OpenAI and Microsoft in 2023 over the use of copyrighted work in their products' answers to user prompts.[11] The list of real-world harms and near-harms from AI systems continues to pile up in the AI Incident Database and the Organisation for Economic Co-operation and Development's (OECD) AI Incidents Monitor.

Beyond the direct organizational impact, AI harms can impact people and create issues in the broader ecosystem in which investors operate – per the National Institute of Standards and Technology's *AI Risk Management Framework* (NIST AI RMF) (see Figure 1).[12]

FIGURE 1 | **The three areas of potential harm posed by AI**

People

Individuals
Harm to civil liberties or human rights, monetary loss

Groups/communities
Discrimination against a population sub-group

Societal
Harm to democratic participation or access to education

Organizations

Harm to business operations

Harm from security breaches or monetary loss

Reputational harm

Ecosystems

Harm to the global financial system, supply chains or other interrelated systems

Harm to natural resources, the environment and the planet

Source: National Institute of Standards and Technology. (2023). *AI Risk Management Framework*.

> **Business leaders overwhelmingly agree that the potential long-term benefits and cost savings associated with implementing AI responsibly far outweigh the initial expense of implementing RAI.**

In recognition of these harms, major regulations related to the oversight of AI are in development, and many are likely to take effect in 2024.[13] Since 2016, countries have passed at least 148 AI-related bills, with the majority passed in recent years.[14] The European Union AI Act (EU AIA), which currently represents the most stringent AI regulation,[15] aims to create a risk-based approach to governing the technology.[16] The EU AIA's extraterritorial scope and its exceptionally high fines (i.e. the higher of up to €35 million or 7% of the company's total worldwide annual turnover) are expected to also influence regulatory requirements outside of the EU's borders (i.e. the "Brussels effect").[17]

While RAI is primarily about mitigating the AI risks and harms above, it can also generate additional benefits for companies. Other sources of value from RAI are predicated upon trust – they include its ability to improve top- and bottom-line growth by increasing customer engagement, broadening revenue streams, offering procurement advantages in competitive bidding processes, and increasing pricing power in the marketplace.[18] Business leaders overwhelmingly agree that the potential long-term benefits and cost savings associated with implementing AI responsibly far outweigh the initial expense of implementing RAI.[19] According to Bain & Company, companies with a comprehensive, responsible approach to AI earn twice as much profit from their AI efforts.[20]

Moreover, companies that start with strong AI governance structures can scale with confidence and drive business growth and innovation.[21] A study by Boston Consulting Group (BCG) found that companies that prioritize responsibility in expanding their AI capabilities experience nearly 30% fewer AI failures[22] compared to companies that do not.[23] Furthermore, companies with mature RAI programmes have double the chance of realizing a business "benefit" – such as better products or services or accelerated innovation – from their AI investment, regardless of their AI maturity.[24]

Further, consumer trust has been identified as a key determinant of corporate performance, explaining 31% of the variance for profit margins and 21% of the variance for return on assets.[25] As AI proliferates, failure to align with RAI principles has the potential to place a drag on this measure. Conversely, bolstering trust through RAI can unlock a flywheel of benefits for governments, businesses and investors.

1.2 | The role of stakeholders in promoting responsible AI

A global consensus has started forming that RAI solutions must keep pace with innovation in AI. As AI's prevalence and potential impact increase globally, companies, investors, governments and regulators will face a potential rise in the scale of risks and consequences.

For investors, building a resilient investment portfolio will involve staying ahead of AI developments. The motivation for investors to accelerate the adoption of RAI is grounded in the anticipation that the worldwide community will persist in moving towards the acquisition, development and implementation of AI, along with the expectation of government-imposed regulations, responding to corporate promises, and evolving consumer and corporate practices.

The _Decision Points in AI Governance_ white paper series from UC Berkeley's Center for Long-Term Cybersecurity details three case studies on efforts to operationalize AI principles.[26]

TABLE 2 | Stakeholders' role and incentive to accelerate the adoption of RAI

Stakeholder grouping	Stakeholders	Examples	Role in RAI	Incentive to accelerate the adoption of RAI
Broader ecosystem	Governments, regulators	Governmental bodies (e.g. EU Commission, US Federal Trade Commission)	Define legal frameworks, enforce AI regulations, set "rules of the game"	Protect citizens, boost innovation, build trust
	Professional and research organizations	Independent organizations, not-for-profits, research institutions (e.g. Partnership on AI, The Alan Turing Institute), academia	Evaluate AI impacts, advance RAI research, advocate for policy	Build trust, influence and community
Investors	Asset owners	Individuals/institutions owning financial assets (e.g. pension funds, insurance companies, sovereign wealth funds, foundations)	Set RAI expectations for asset managers (internal and external), contribute to standard setting	Manage risks, create value, act in best interests of their stakeholders (fiduciary duty)
	Asset managers	Individuals/institutions managing investments (e.g. private equity general partners, hedge fund investment managers, asset owner employees)	Integrate RAI into investment decisions and engagement with companies	Manage risks, create value, deliver on asset owner mandates
Companies	Boards	Governing bodies within companies (e.g. board of directors)	Oversee RAI strategy, manage AI risks	Protect reputation, comply with regulations, direct strategy, minimize harm to people and planet
	Management	Teams that execute company operations (e.g. C-suite executives, operations managers and other employees)	Create RAI governance, implement RAI strategy, ensure operational RAI alignment, minimize harm to people and the planet	Build a sustainable business, gain a competitive edge, advance best practices for product development

2 | Investor engagement in RAI

To mitigate risks and harms, investors will need to engage on RAI across the portfolio and potentially beyond.

2.1 | Focusing on areas of greatest impact

❝ Investors should adopt an AI governance framework with clear policies and standards and promote a culture of accountability and transparency around AI development, adoption and use.

Where can investors begin? The following steps will help guide them as they begin to engage with companies, external asset managers and the broader ecosystem. Additional tools have also emerged to assist them.

Step 1: Develop RAI commitments and apply its principles and practices to internal operations.

The first step for investors looking to integrate RAI across their portfolios is to become knowledgeable on AI/RAI and apply it to their own operations. This includes defining their own RAI principles and priorities, including what they will **not** invest in. If using AI in their own operations, investors should adopt an AI governance framework with clear policies and standards and promote a culture of accountability and transparency around AI development, adoption and use.

Step 2: Conduct RAI due diligence on the portfolio.

AI will soon be at play in nearly all companies, from the innovation labs of big tech to factory floors. The question for investors is, where in the portfolio should they first focus on RAI? Given the diversity of portfolios and investor archetypes, engagement and mitigation strategies should be tailored to individual portfolios and investment strategies.

Investors should perform proper due diligence to assess how companies and investment partners are positioned to meet RAI principles (see case study 1). In portfolios of direct investments, potential areas for strengthening or instilling RAI practices may occur where AI drives core revenue streams, is deployed in high-risk areas, is highly regulated, or has high adoption rates (see case study 2). In addition, companies in the "real economy" that have not yet adopted AI in any meaningful way represent an opportunity for investors to help make RAI a part of the company's AI strategy foundation. In portfolios with outsourced asset management, investors may perform diligence on their investment

partners to ensure appropriate RAI policies and procedures are in place.

Step 3: Engage with companies, external managers and the broader ecosystem.

Working with companies, external managers and other players can extend investors' influence and help them maximize the value of their AI investments by building bespoke plans to accelerate the adoption of RAI. Engagement with companies, external managers and the broader ecosystem is covered more in the following chapters of this playbook.[27]

Investor priorities will often depend on what is discovered during due diligence and stakeholder needs (e.g. corporate governance, capacity building, strategy and risk management). That said, **AI governance** is a key point of leverage and impact.

1. **Governance establishes the foundation**. Mechanisms like regulation, principles, policies, standards and oversight lay the groundwork for RAI.

2. **There is a gap before regulation**. Laws and regulatory expectations governing AI are still emerging in most regions. This creates an opportunity for investors to promote strong self-governance mechanisms that proactively address stakeholder concerns. Getting ahead of regulation helps proactively mitigate regulatory and business risks, saves costs and ultimately influences the regulatory landscape.

3. **Governance supports innovation and trust**. With effective governance in place, organizations can confidently pursue innovative applications of AI, knowing they have mechanisms to identify and mitigate risks.

Overall, ecosystem AI governance informs and enforces organizational AI governance, which leads to the delivery of RAI principles. The transparency provided to investors allows them to hold corporates and investment partners accountable.

Radical Ventures' RAI for start-ups framework

Radical Ventures, an AI-focused venture fund, developed a framework to help investors assess the efforts of early-stage companies as they build and deploy AI.

Investors can use Radical Ventures' Responsible AI for Start-ups (RAIS) framework, available as an open-source tool, early in the investment process to assess whether start-ups are maximizing AI's potential benefits and building appropriate risk mitigation measures.

RAIS assigns a score for each vulnerability or threat uncovered, based on the likelihood of that threat occurring and the magnitude of the consequence. When safety risks are identified in RAIS, the framework provides mitigation strategies. It helps investors to think through best practices (such as those documented in critical algorithm studies) and to build those practices into their core business strategy. The framework covers three key areas of risks: social and ethical, regulatory compliance, and technical.

While tailored to venture capital (VC) investors that may be involved with helping founders develop products at an early stage, the RAIS framework is relevant for anyone engaging with AI tools. It's designed to help users answer three questions: What are the benefits? What are the risks? Do the benefits outweigh the risks?

Other diligence tools

The Data & Trust Alliance, Responsible Data & AI Diligence for M&A

This toolkit helps investors assess the value and risks of business models built on data and AI. Despite the reference to mergers and acquisitions (M&A) in the title, the tool can be tailored to investors seeking minority positions.

TechBetter, *Responsible Investing in AI: A Guidebook for VCs*

TechBetter's guidebook offers a five-step process: 1) assessing whether a company poses AI risks, 2) rating the company's regulatory risk and risk of conflict with the investors' values, 3) assessing the company's RAI maturity, 4) determining investment eligibility, and 5) defining post-investment support.

United Nations Human Rights Office of the High Commissioner, B-Tech initiative

This tool provides guidance and resources for implementing the United Nations Guiding Principles (UNGPs) on business and human rights in technology.[28] It aims to help institutional investors engage with technology companies on human rights risks linked to core elements of their business models.

Insights from Credo AI on sectoral opportunities for RAI

Credo AI, an AI governance platform, works with enterprises across sectors to responsibly build, adopt, procure and use AI at scale. Credo AI has found that the sectors offering the greatest opportunity for RAI share several characteristics: the use of AI could raise material business risks, AI tools are already widely used, and the use of AI could have significant financial or reputational impact.

FIGURE 2   **AI risk and adoption by sector**

High risk   Low adoption
– Consumer
– Discretionary

High risk   High adoption
– Healthcare
– Communication services
– Financials
– Industrial

Low risk   Low adoption
– Consumer staples
– Materials
– Real estate
– Utilities

Low risk   High adoption
– IT
– Energy

Risk of AI systems

Adoption of AI systems

Credo AI's customer base suggests that companies in sectors like healthcare, communication services, financials and industrials – encompassing services in human resources (HR) and transport – are more inclined to invest in and adopt RAI practices due to the potential impact on individuals (see Figure 2). Misuse of AI in these sectors could have serious consequences, including health risks, financial loss or even loss of life. These sectors are also facing growing regulatory pressures and the potential for public scrutiny.

Additionally, there are sectors where the adoption of AI is high, but the perceived risks of their use cases are relatively low, especially in terms of direct harm to individuals or regulatory non-compliance. These include information technology (IT), where AI is used to optimize processes, enhance operational efficiency or provide non-critical consumer services. Here, the emphasis on RAI often dovetails with more traditional considerations when adopting digital technologies at scale, such as privacy, data protection and security.

Cross-sector AI use cases

AI applications, including employment decision tools designed to select potential employees and content generation tools like Microsoft Copilot, span multiple industries. For example, the rising adoption of generative AI tools underscores the universal importance of RAI principles, regardless of industry-specific risks or AI adoption levels.

2.2 | Engagement with companies

A company's investors, board directors and management all play critical roles in creating sustained long-term value. The division of authority and responsibilities among this triad of roles is the core of good governance.[29]

In respecting this division of authority and responsibilities, investors engage with portfolio companies through their board representation and shareholder voting rights.[30] Directors of companies are responsible for providing oversight and counsel to management about material business opportunities and risks, and ensuring that these are considered in strategy-setting, operational implementation and in relevant disclosures

to the market. Where AI is a material risk or opportunity, the board should assume oversight as it would any other material business risk.

Productive engagement with the boards of investee companies is critical. The International Corporate Governance Network (ICGN), led by investors responsible for assets under management of around $70 trillion, has published an AI engagement guide for investors,[31] inclusive of key considerations, expectations of companies and stewardship dialogue. Table 3 highlights sample objectives, tools and examples to further help investors strategize engagement with boards on RAI.

TABLE 3 | **Productive engagement with boards on RAI**

Issue	Engagement objectives	Supporting tools	Examples
AI principles	Adoption and publication of RAI principles	*NIST AI RMF*, *OECD AI Principles*, *EU AI Act*, *Blueprint for an AI Bill of Rights*	Telefonica, BMW, Novartis, Allianz SE and H&M Group have all published AI principles The World Benchmarking Alliance Collective Impact Coalition for Digital Inclusion (WBA Digital CIC), led by Boston Common Asset Management and Fidelity International, operates a coordinated engagement campaign to push technology companies to advance ethical AI policies and practices. Of the 200 companies evaluated in the 2023 Digital Inclusion Benchmark 19 have since announced their inaugural AI principles[32] – Digital CIC members had led outreach to all 19 throughout the year
Policies and procedures	Alignment with emerging global RAI frameworks, standards and certifications Independently audit AI systems, guidelines and risk management processes	AI governance frameworks such as the *NIST AI RMF*, *Model AI Governance Framework – Personal Data Protection Commission (PDPC) Singapore* ISO/IEC's standards for AI (e.g. ISO/IEC JTC 1/SC 42, ISO/IEC 42001) RAI Institute's Certification Program for AI Systems AI Verify validates the performance of AI systems, excluding generative AI/large language models (LLMs); DecodingTrust evaluates LLMs Open Worldwide Application Security Project's (OWASP) *LLM AI Cybersecurity and Governance Checklist*	Microsoft and IBM have developed policies and procedures that seek to align with emerging global RAI standards Radical Ventures includes a clause in its term sheets asking for a commitment from investee companies to work with Radical to develop and implement RAI guidelines
Roles and responsibilities	Demonstrable AI competence and oversight Clear definitions of roles and responsibilities for various AI functions Inclusion of AI/RAI on board agendas	The Open Compliance and Ethics Group (OCEG) provides guidance on key questions boards should be asking[33] The National Association of Corporate Directors' (NACD) *Director Essentials: Digital Transformation: AI and Board Governance*	Court Square Capital Partners partnered with Accenture to host a virtual AI tech-in for their portfolio companies[34] UBS Asset Management engaged with a company integrating a proprietary LLM into all its products and services, focusing on board capability and accountability on AI governance, and the scope of the company's AI committee interactions with executives leading AI development

TABLE 3 | Productive engagement with boards on RAI (continued)

Issue	Engagement objectives	Supporting tools	Examples
Investor transparency	Company disclosure on RAI alignment Public commitments to RAI Transparency over data and how models are trained	Potential concepts include model cards, datasheets, dashboards and transparency notes, among others Metrics around the governance process, including whether the company has an inventory of AI use cases, risk assessments, risk mitigation strategies and ongoing monitoring Broader sustainability reporting frameworks like the Global Reporting Initiative (GRI) and the International Financial Reporting Standards (IFRS) could include AI governance as they evolve (although the IFRS S1 can be used in its current form)	Microsoft, IBM, Google and Meta disclose their AI efforts and progress. Salesforce and Google are using model cards (e.g. Gemma model card) Telus has made commitments to RAI.[35] It is one of the first signatories to Canada's Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems *The Interfaith Center on Corporate Responsibility (ICCR) 2024 Proxy Resolutions and Voting Guide* highlights 11 proposals filed for the 2024 proxy season, illustrating growing investor concern about AI's potential impact. For example, resolutions were filed against Alphabet related to its AI principles and board oversight, reporting on generative AI misinformation/disinformation risks, and human rights impact of AI driven ad practices

CASE STUDY 3
How NBIM makes RAI a priority

Norges Bank Investment Management (NBIM), manager of the $1.5 trillion Norwegian Government Pension Fund Global, published its view on RAI in 2023.[36] NBIM's view is rooted in the understanding that as AI becomes more integrated into the operations of companies across sectors and markets, its use must be aligned with responsible business conduct, which includes ethical considerations, human rights and shareholder value. By enhancing innovation and efficiency, AI can result in significant gains for companies, individuals and society. However, AI is developing rapidly, making it challenging to predict and manage risks.

NBIM views responsible development and use of AI as a core element of responsible business conduct and a necessary complement to the emerging regulatory landscape. NBIM emphasizes the following three elements of RAI:

1. **Board accountability:** NBIM believes the board of directors is accountable for ensuring that companies develop and use AI responsibly. NBIM expects boards to have suitable competency and to exercise appropriate oversight of AI strategies, ensuring they are integrated into broader risk management frameworks and align with international norms and standards.

2. **Explainability and transparency:** Understanding how AI models are trained and make decisions is important for ensuring AI-driven decisions are fair, unbiased and can be trusted by stakeholders. NBIM encourages companies to be transparent about the datasets used for training algorithms and the measures taken to prevent discriminatory outcomes.

3. **Robust risk management:** NBIM believes that companies must be proactive in managing AI-related risks and transparent about their objectives in developing and deploying AI systems. Risk management includes identifying, assessing and mitigating the potential and actual adverse impacts of AI. Companies should exercise human oversight and control, secure all AI systems to prevent malicious use and data breaches and understand the wider societal impacts associated with the deployment of AI, such as long-term effects on the workplace as well as inequality, privacy and discrimination.

NBIM engages with individual companies, particularly those in the healthcare and technology sectors, through dialogue and exercising its shareholder rights. NBIM has sent letters to over 60 company boards to share its views and encourage the adoption of RAI practices.

UBS Asset Management's model for board engagement on RAI

As AI integration moves beyond concepts and pilots into real business operations, UBS Asset Management (UBS AM) believes it is important to strengthen governance and internal controls.

UBS AM focuses on AI governance, particularly how boards and executive management teams prioritize RAI, stressing the importance of internal accountability and communication across business functions to support board governance throughout the implementation cycle.

UBS AM expects boards to determine if the right business functions are involved in the AI life cycle. From UBS AM's engagement experience, the "missing link" often lies in connecting feedback from employees and customers to the group layer to improve the overall structure, policies, standards and AI inventory and to the business line for version controls and customization.

Table 4 spans three layers of the organization and three categories of risk, with sample activities at each level and groups to involve. This framework translates conceptual AI governance models (e.g. data, model and application layers) into a corporate governance model.

Flexibility is required when applying the framework to individual organizations. While this framework was designed for large multi-layered organizations, the conceptual model can be applied to smaller organizations.

TABLE 4 **UBS AM's three-layer AI oversight model**

Layer	Sure risks "known knowns"	Expected risks "known unknowns"	Unexpected risks "unknown unknowns"
Group level Sets the stage for AI across the organization. Includes adopting AI principles, governance structures, clear policies and ensuring standards are met	Requires robust governance framework and principles, comprehensive AI policies and standards, and centralized technology functions ☐ Legal and compliance ☐ IT ☐ Risk ☐ Data	Strategic risk management processes, learning and improvement cycles are established for risk preparedness ☐ Legal and compliance ☐ Risk ☐ Audit ☐ Data	A culture of agility, integrity and resilience is developed. Strong external partnerships are maintained for rapid response and recovery ☐ Legal and compliance ☐ Risk ☐ Audit
Business line Applies AI governance in departments (e.g. HR, finance) or business units. Aligns AI initiatives with broader organizational standards, managing risks through assessments and prepares for contingencies	Ensures AI initiatives align with organization-level governance, compliance and risk management within specific business units ☐ Legal and compliance ☐ Sustainability ☐ Department/business line	Risk assessments and impact analyses, contingency plans are carried out for critical AI systems ☐ Legal and compliance ☐ Risk ☐ Audit ☐ Finance ☐ Sustainability ☐ Department/business line	Business continuity planning, emergency procedures and crisis management plans are established ☐ IT ☐ Risk ☐ Audit ☐ Department/business line
Application and use case Ensures each AI system adheres to established standards, monitors for new risks and maintains incident response plans	Requires adherence to development and deployment standards, conduct thorough testing and validation ☐ Finance ☐ Sustainability ☐ Business line ☐ Operations ☐ Other enabling functions	AI systems are monitored for emerging risks and performance issues, mechanisms for real-time risk detection and mitigation ☐ IT ☐ Risk ☐ Audit ☐ Business line ☐ Operations	Rapid incident response protocols, and swift rollback or shutdown procedures for AI systems for unforeseen issues are in place ☐ Legal and compliance ☐ Risk ☐ Operations ☐ Other enabling functions ☐ Department/business line

Note: For illustration purposes only.

Source: UBS AM

2.3 | Engagement with external asset managers

External asset managers are essential to the effort to apply RAI in a consistent manner across the portfolio. By engaging them, asset owners can help ensure that RAI is implemented in alignment with the investor's values and principles.

An important first step in this process is to agree on goals and align incentives.

The following is a list of questions to help begin that discussion:

BOX 3 | **Asset manager discussion guide**

– How does the asset manager ensure they have sufficient understanding and expertise on the topic of AI, specifically in relation to risk management and value creation?

– Has the concept of RAI been factored into the asset manager's investment strategy? If so, how? If not, why not?

– Has the asset manager expressed a stance on RAI? Is this reflected in existing policies such as their investment policy?

– Can the asset manager outline the risk management processes in place to identify

and mitigate material AI-related risks? From the asset manager's perspective, what are the key AI-related risks for investment activities?

– Does the asset manager conduct risk-based due diligence to identify, prevent and mitigate any negative impacts of AI in the portfolio?

– How does the asset manager ensure the companies they invest in are adhering to RAI practices?

– What measures or key performance indicators does the asset manager use to assess RAI postures across the portfolio?

Table 5 covers some sample objectives, tools and examples investors can consider when engaging asset managers on RAI in their portfolios (asset managers' own use of AI would be covered per section 2.2).

It is noted that there are few public examples of asset owners engaging with external asset managers on RAI. That said, there is an opportunity for asset owners to communicate their positions on RAI to their external asset managers.

TABLE 5 | Creating productive engagement with asset managers on RAI

Issue	Engagement objectives	Supporting tools	Examples
Policies and procedures	Integration of RAI into diligence and investment decisions	Limited partner (LP) signatories of Responsible Innovation Labs' (RIL) RAI commitments make pledges that include encouraging their VC general partners to make voluntary RAI commitments. RIL has created an RAI Protocol to help put the commitments into practice. The ICGN - GISD Model Mandate provides guidance to asset owners to ensure stewardship objectives are fully reflected in investment management agreements and contract terms.[37]	Princeton University Investment Company and Oxford University Endowment Management encourage their VC general partners to make RIL's voluntary RAI commitments.[38]
Investor transparency	Reporting on portfolio alignment with RAI, including performance metrics, risk assessments and other considerations	Industry-standard due diligence questionnaires from the Institutional Limited Partner Association (ILPA) and the Alternative Investment Management Association (AIMA) can be expanded to address questions related to AI and RAI and provide guidance on disclosure.	While examples could not be found as of the date of this report, this type of reporting can be expected to become more common.

CASE STUDY 5
External asset manager engagement at Manulife[39]

Manulife Investment Management (Manulife IM) is the global wealth and asset management segment of Manulife Financial Corporation. In 2023, Manulife IM Private Equity & Credit (PE&C), a group within Manulife IM, initiated a research project focused on building knowledge on topics related to responsible technology innovation and implementation (RTII), also referred to as sustainable digital economy (SDE).

As long-term investors and fiduciaries, Manulife IM PE&C prioritizes downside protection and long-term capital growth for clients, emphasizing the importance of understanding the technological factors that influence future business success. Manulife IM PE&C acknowledges the profound effects of AI, automation and digitalization in various industries and aims to integrate responsible governance of RTII considerations into its investment process, seeking to further enhance the resilience and long-term returns of its investment portfolios.

To date, Manulife IM PE&C has engaged with an external research collective, Investors for a Sustainable Digital Economy (ISDE), contributing to the creation of a sector impacts map for RTII.[40] Manulife IM PE&C has participated in roundtables on the future of work and surveillance and contributed to a research report titled *Bits and Bots: How Digitalization & Automation is Reshaping the Workforce & Work Itself*. Internally, the Manulife IM PE&C group formed a working group to incorporate RTII research into its investment process and included the SDE theme in its most recent annual sustainable investing questionnaire sent out to general partners (GPs), sponsors and portfolio companies.

Planned actions relating to SDE for 2024 include continued refinement of questionnaires to better understand RTII practices, engaging with select GPs to deepen Manulife IM PE&C's knowledge and evolve its sustainable investing tools and frameworks. This will assist in the effort to identify gaps and opportunities to further advance its approach to manager/sponsor evaluations.

Looking ahead, Manulife IM PE&C anticipates RTII will become increasingly important for value creation, requiring investors and businesses to develop and integrate new practices and strategies.

2.4 | Engagement with the broader ecosystem

To create a structured approach for elevating RAI standards, investors must work collectively with an array of stakeholders. These include investment partners, external managers, companies, government and regulators, professional and research organizations, and other actors and observers. The ESG Data Convergence Initiative (EDCI) is an example of how large investors can influence the broader ecosystem to adhere to proposed governance standards. EDCI was formed in 2021 by a group of influential GPs and limited partners (LPs) to standardize performance-based environmental, social and governance (ESG) metrics. The effort by more than 400 GP and LP members used data from 4,300 portfolio companies and resulted in a benchmarking software platform.

Table 6 provides sample objectives, tools and real-world examples investors can consider when engaging on AI across the ecosystem:

TABLE 6 | **Engaging with the ecosystem to encourage RAI**

Issue	Engagement objectives	Supporting tools	Examples
Regulatory and legal	Input on investor viewpoints on policy or directly to policy-makers Creation of clear, enforceable standards for RAI	*A Sustainable Finance Policy Engagement Handbook* published by the United Nations Principles for Responsible Investment (UN PRI)	149 global institutional investors representing over $1.66 trillion in assets signed an Investor Statement on the EU's AI Act, calling for strong regulation while acknowledging the potential benefits of AI.
Voluntary governance frameworks	Contributions to the development of principles, policies, guidelines, standards and certifications	Coalitions or networks of investors and investment partners to share best practices, develop RAI standards and leverage their collective bargaining power, such as the PRI Collaboration Platform	Candriam champions an initiative examining the ethics of facial recognition technology and warning of its ethical hazards.[41] Candriam leads a group of 55 investors managing more than $5 trillion in assets. The group is calling for an ethical approach to facial recognition technology, citing concerns about its potential for misuse.[42]
Investor transparency	Creation of reporting and disclosure standards Ecosystem understanding of investor expectations on AI/RAI Disclosure of how investors are incorporating AI/RAI into their investment decisions	The ISSB Investor Advisory Group (IIAG) is a group of leading asset owners and managers in various markets committed to improving the quality and comparability of sustainability-related financial disclosures	Norges has made public its views on responsible artificial intelligence.[43] General Catalyst has published *A Manifesto for Responsible AI*. CPP Investments identified responsible sourcing and deployment of AI as a sustainable investing principle in its *Policy on Sustainable Investing*. Federated Hermes published a white paper, *Investors' expectations on responsible artificial intelligence and data governance*, and an AI assessment framework outlining the importance of AI as a sustainability consideration for investors. HSBC Asset Management co-authored *Investors' Expectations of Ethical Artificial Intelligence in Human Capital Management*.

Temasek's approach to promoting RAI

Early on, Temasek recognized AI as a transformative force driving industry evolution and global change. In 2018, Temasek established the AI Pod, a team dedicated to seizing AI-related business opportunities while promoting RAI adoption. Temasek takes a broad-based approach in its efforts to drive RAI practices by collaborating with regulators, industry associations and other entities.

Engagement with regulators and industry associations:
Deputy Chief Executive Officer Chia Song Hwee's involvement in the Infocomm Media Development Authority/Personal Data Protection Commission's (IMDA/PDPC) Advisory Council on the Ethical Use of AI and Data underscores Temasek's commitment to AI ethics and governance issues. The council formulates ethical guidelines and policies that govern the use of AI and data-driven technologies in the private sector. By providing expert insights and recommendations, the council aims to ensure that AI deployments adhere to ethical principles and promote societal well-being.

Partnerships with industry associations:
Temasek's AI Pod joined the Singapore Computer Society's (SCS) AI Ethics & Governance Professional Certification Committee in developing one of the first training and certification programmes that promotes ethical AI practices.

By providing comprehensive training and certification, the committee aims to equip individuals with the knowledge and skills to navigate ethical challenges in AI development and deployment.

Venture building investments:
Temasek has made select venture building investments in Resaro and Aicadium. Resaro offers independent, third-party assurance solutions for AI systems to enhance trust in AI and accelerate responsible, safe and robust AI adoption. Aicadium collaborates with Temasek's portfolio companies to innovate and scale AI products aligned with responsible AI principles, to improve business outcomes while helping people thrive and supporting sustainability.

Ecosystem engagement:
Temasek actively supports the AI Verify Foundation, a global open-source community that convenes AI owners, solution providers, users and policy-makers to build trustworthy AI. The AI Verify Foundation aims to enhance trust in AI systems by providing standardized tests and promoting transparency in AI performance. By mitigating risks associated with AI, such as biases and vulnerabilities, AI Verify facilitates a safer, fairer and more robust AI ecosystem.

The hurdles ahead – and how to clear them

Strategies for effective RAI engagement are still emerging.

> **Investors, academics and non-governmental organizations must bridge differing goals, jargon and trust issues to collaborate effectively and raise awareness of RAI practices.**

Investors face a range of challenges in their efforts to engage on RAI, from grappling with regulatory uncertainty to navigating the rapidly evolving landscape of AI development and use. There is still much work to be done, including:

Establishing a dynamic framework for AI governance: Investors need an approach to engaging on RAI that can adapt as AI evolves and is sophisticated enough to reflect the various technologies under the AI umbrella. Developing agile governance practices can help organizations anticipate and adapt to changes in technology, regulation and societal impacts. Creating reporting mechanisms that are not just tick-box exercises but provide meaningful insight into AI initiatives requires careful design and regular review.

However, it's important to note that AI issues are highly domain-dependent, with issues of fairness in credit scoring and healthcare diagnostics, for instance, being vastly different. Therefore, assessing the risk and impact can help generate a rank-ordered list of the most important domains and applications an organization should tackle. This list can then be adapted to changes in technology or the context of its use.

Development of standardized RAI metrics: Many authoritative sources of investor guidance on metrics and disclosure (e.g. MSCI and the International Sustainability Standards Board) have yet to specifically address AI, and high-quality metrics are only just emerging. The OECD's Artificial Intelligence Policy Observatory has begun cataloguing metrics for AI trustworthiness and AI risks.[44] Stanford University's Holistic Evaluation of Language Models (HELM) project lists 59 metrics for large language model evaluation.[45] Finding metrics that can work with different AI use cases is a challenge, although evaluating the oversight process itself (i.e. process metrics) is a scalable approach that can provide a level of assurance while this area matures.

Capacity building: The fast-paced development of AI demands continuous learning and adaptation by investors. Asset owners may struggle to set clear expectations for asset managers and corporations. Directors may lack the expertise to comfortably oversee AI risk and opportunity.

Yet, as with cybersecurity, the Securities and Exchange Commission (SEC) expects boards to fulfil their supervisory duties in technical areas that pose significant risk to the company. Some organizations in regulated industries have steered clear of direct customer-facing applications of generative AI until they better understand the technology.[46]

Several investor-led initiatives are under way to speed capacity building. In the public markets, Boston Common Asset Management and Fidelity International lead a World Benchmarking Alliance coalition that aims to raise awareness of the importance of responsible and ethical AI, disseminate best practices and improve technology companies' commitment to ethical AI. Temasek's AI Pod joined the Singapore Computer Society's (SCS) AI Ethics & Governance Professional Certification Committee in developing one of the first training and certification programmes.

Improved guidance on the financial materiality of RAI: While a body of knowledge continues to build around the financial materiality of RAI, recognition of RAI's business case as a driver of value creation is not acknowledged in all circles. Demonstrating the long-term benefits and potential for competitive advantage through RAI practices can help overcome these barriers.

A study by IBM and Notre Dame University proposes a framework to evaluate the Return on Investment in AI Ethics.[47] It identifies three types of returns, including direct economic impact (cost savings, revenue generation or reduction in cost of capital), intangible impact (indirect returns that build reputation with stakeholders and lead to economic returns over time), and real options and capabilities (organizational capabilities in areas like risk assessment, regulatory compliance and software tools) that provide future flexibility and cost savings.

Stakeholder alignment and change management: Aligning interests and strategies across diverse investors and partners can be challenging, requiring effective communication and compromise. Investors, academics and non-governmental organizations (NGOs) must bridge differing goals, jargon and trust issues to collaborate effectively and raise awareness of RAI practices.

Asset managers might have reservations due to concerns about disruption, added burden and perceived value. Engaging with the ecosystem presents investors with unique challenges, including the complexity of political and regulatory landscapes and the often lengthy process of reaching a consensus on regulations. It is important for investors to find a balance between advocacy and adaptability, adjusting to changing environments.

Global laws and regulation: There is a need for clarity and certainty around prospective regulatory regimes and applications of the law. Investors and companies seek guidelines and frameworks to navigate the complex landscape. Such clarity would enable more confident and consistent engagement on RAI.

That said, waiting for clarity is not an excuse to delay the RAI journey. Companies and investment partners can benefit when investors, at the very least, communicate the minimum standards or thresholds they expect.

Navigating labels: While AI and RAI could become a component of ESG, many investors see these concepts as distinct. What's more, the politicization of ESG has complicated any potential

association of the term with efforts to advance RAI (e.g. Bluebell Capital Partners' opposition to BlackRock's ESG efforts[48] and the politicization and retreat from diversity, equity and inclusion).[49] Regardless of labels, AI remains an emerging business risk and opportunity that will determine the long-term sustainability of enterprises. As such, boards should oversee the incorporation of RAI into management's strategic planning and operational execution. In addition, they should insist on market disclosure where AI adoption, development and use is material. Regulatory protections to make these decisions free of political backlash would also be beneficial.

Tension between RAI and corporate imperatives: The tension between RAI and corporate imperatives (like speed to market) extends to investors. It is important that investors acknowledge this tension and explore ways to address it, for example, by shifting expectations around short- versus long-term value creation.

FCLTGlobal develops research and tools that encourage long-term investing. Similarly, investors with certainty of capital (such as pensions) are better able to weather short-term volatility and take a long-term approach.

Conclusion

The transformative potential of AI is undeniable, and with this potential comes a compelling economic rationale for prioritizing RAI. While challenges such as navigating technical complexities, adapting to regulatory changes and establishing consensus on governance are present, they are surmountable with a concerted and collective effort from all stakeholders.

Investors play a pivotal role in this endeavour. By championing RAI, they not only help mitigate risks but also unlock opportunities for sustainable growth and innovation. This playbook serves as a strategic guide, offering insights and actionable strategies for investors to drive the integration of RAI in AI development and deployment.

For investors looking towards the horizon of long-term value creation, advancing RAI is a strategic business decision. Through collaboration, education and proactive engagement, investors can help steer the AI revolution towards a future where technology operates with responsibility at its core, benefiting companies, consumers and society at large.

Contributors

Lead authors

Chris Gillam
Director, Operational Risk, CPP Investments;
Responsible AI Playbook for Investors Project
Fellow, World Economic Forum

Devendra Jain
Lead, Artificial Intelligence, Quantum Technologies,
World Economic Forum

Acknowledgements

Community members

Travis Antoniono
Investment Director, CalPERS

Jim Breyer
Founder and Chief Executive Officer, Breyer Capital

Michael Brent
Director, Responsible AI, Boston Consulting Group

Cansu Canca
Director, Responsible AI Practice, The Institute
for Experiential AI, Northeastern University

Navin Chaddha
Managing Partner, Mayfield Fund

Benjamin Chekroun
ESG Analyst, Candriam

Christine Chow
Managing Director, UBS Asset Management

Caroline Eriksen
Head, Social Team, Active Ownership,
Norges Bank Investments Management

Theodoros Evgeniou
Professor of Decision Sciences and
Technology Management, INSEAD

Brian Franz
Global Chief Information Officer and
Head of Enterprise Resiliency, State Street

John Hearty
Vice-President, AI Governance, Mastercard

Derek Ip
Director, Sustainable and Impact Investing,
UBS Asset Management

Matthew Katz
Senior Managing Director; Global Head
of Data Science, Blackstone

Tony Kim
Managing Director; Head, Technology Investing,
Fundamental Active Equity, BlackRock

Christopher Kauffman
Principal, General Catalyst

Alyssa Lefaivre Škopac
Head, Global Partnership and Growth,
Responsible AI Institute

Giovanni Leoni
Head, Business Strategy and Development, Credo.AI

Art Levy
Chief Business Officer, Brex

Leah Morris
Senior Director, Velocity Program, Radical Ventures

Greg Nielsen
Senior Managing Director, Private Capital,
Ontario Teachers' Pension Plan

Michael Ortiz
Senior Director, Policy, Sequoia Capital

Nick Pelosi
Associate Director, Engagement, Federated Hermes

Philipp Raether
Group Chief Privacy Officer, Allianz

Endnotes

1. Schwab, K. (2016). *The Fourth Industrial Revolution: What it means, how to respond*. World Economic Forum. https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/.

2. See for example: Evgeniou, T. (2023). *A Very Brief Technical History of AI. LinkedIn.* https://www.linkedin.com/pulse/part-i-very-brief-technical-history-ai-theodoros-evgeniou-tipfe/.

3. Anyoha, R. (2017). *The history of artificial intelligence*. Harvard. https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/.

4. The report does not provide timelines for capturing impact. McKinsey & Company state that how quickly impact is achieved is determined by how quickly and effectively management teams adopt use cases and the necessary complementary management changes.

5. Chui, M., et al. (2023). *The Economic Potential of Generative AI: The Next Productivity Frontier*. McKinsey & Company. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#key-insights.

6. *Responsible AI Institute*. (2024). *Responsible AI*. https://www.responsible.ai/faq/.

7. National Institute of Standards and Technology (NIST). (2023). *Artificial intelligence risk management framework (AI RMF 1.0)*. https://doi.org/10.6028/NIST.AI.100-1.

8. Renieris, E., et. al. (2022). *To Be a Responsible AI Leader, Focus on Being Responsible*. MITSloan Management Review. https://sloanreview.mit.edu/projects/to-be-a-responsible-ai-leader-focus-on-being-responsible/#:~:text=Notably%2C%2042%25%20of%20our%20respondents,lag%20behind%20strategic%20AI%20priorities.

9. Cheatham, B., et al. (2019). *Confronting the risks of artificial intelligence*. McKinsey Global Institute. https://www.mckinsey.com/capabilities/quantumblack/our-insights/confronting-the-risks-of-artificial-intelligence.

10. *The Wall Street Journal. (2023). Zillow's shuttered home-flipping business lost $881 million in 2021.* https://www.wsj.com/articles/zillows-shuttered-home-flipping-business-lost-881-million-in-2021-11644529656.

11. *The New York Times*. (2023). *The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work*. https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html.

12. National Institute of Standards and Technology (NIST). (2023). *Artificial intelligence risk management framework (AI RMF 1.0)*. https://doi.org/10.6028/NIST.AI.100-1.

13. Casovan, A., & Shankar, V. (2022). *"Soft law" to regulate AI: The role of standards, conformity assessments and certification programs. Responsible AI Institute.*

14. Maslej, N., et al. (2024). *The AI index 2024 annual report.* AI Index Steering Committee, Institute for Human-Centered AI, Stanford University.

15. Casovan, A., & Shankar, V. (2022). *"Soft law" to regulate AI: The role of standards, conformity assessments and certification programs. Responsible AI Institute.*

16. European Commission. (n.d.). *A European approach to artificial intelligence*. https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence.

17. *Sidley*. (2022). *Proposal for EU artificial intelligence act passes next level – Where do we stand and what's next?* https://www.sidley.com/en/insights/newsupdates/2022/12/proposal-for-eu-artificial-intelligence-act-passes-next-level.

18. *Economist Intelligence Unit*. (2020). *Staying ahead of the curve – The business case for responsible AI*. https://www.eiu.com/n/staying-ahead-of-the-curve-the-business-case-for-responsible-ai/.

19. Ibid.

20. Sinha, V., et al. (2024). *Adapting your organization for responsible AI*. Bain & Company. https://www.bain.com/insights/adapting-your-organization-for-responsible-ai/.

21. *Credo AI*. (2023). *IDC and Credo AI survey: The business case for responsible AI*. https://www.credo.ai/blog/rai-survey-2023.

22. Failures are defined as an AI system functioning in an unintended way that impacted the company, employees, customers or society. The types of impacts may have included financial, regulatory, legal, reputational or violation of values.

23. *Boston Consulting Group (BCG)*. (2022). *A responsible AI leader does more than avoiding risk*. https://www.bcg.com/en-ca/publications/2022/a-responsible-ai-leader-does-more-than-just-avoiding-risk.

24. Ibid.

25. PricewaterhouseCoopers (PwC). (n.d.). *Translating trust into business reality*. https://www.pwc.com/gx/en/issues/trust/translating-trust-into-business-reality.html.

26. See also this case study (available on request): McElhaney, K., G. Smith, I. Rustagi, O. Groth. (2022). *Responsible A.I.: Tackling Tech's Largest Corporate Governance Challenges. Berkeley-Haas Case Series*. University of California, Berkeley Hass School of Business. https://sk.sagepub.com/cases/responsible-ai-tackling-techs-largest-corporate-governance-challenges.

27. Investor engagement strategies (such as shareholder resolutions, proxy voting, and negotiating terms into investment management agreements) are common across topics and are not unique to RAI. Resources such as the United Nations Principles of Responsible Investment (UN PRI) provide detailed guidance on engagement tactics across different targets and asset classes – this paper does not re-examine them in detail.

28. *Office of the High Commissioner for Human Rights*. (n.d.). *B-Tech Project*. https://www.ohchr.org/en/business-and-human-rights/b-tech-project.

29. *CPP Investments*. (n.d.). *Our Investment Strategy*. https://www.cppinvestments.com/the-fund/how-we-invest/our-investment-strategy/#1600181159050-82b79d73-e840.

30. *CPP Investments*. (2023). *2023 Report on Sustainable Investing.* https://www.cppinvestments.com/wp-content/uploads/2023/10/SI-Report-2023-EN.pdf.

31. International Corporate Governance Network (ICGN). (2020). *Artificial Intelligence and Board Effectiveness.* https://www.icgn.org/sites/default/files/2021-06/ICGN%20Viewpoint%20Artificial%20Intelligence.pdf.

32. World Benchmarking Alliance. (2023). *Augmenting Ethical AI: 2023 Progress Report on the Collective Impact Coalition for Digital Inclusion.* https://assets.worldbenchmarkingalliance.org/app/uploads/2023/09/Digital-CIC-2023-Progress-Report_.pdf.

33. Switzer, C. (n.d.). *Critical Questions About AI That Every Executive Team Should Be Asking*. OCEG. https://www.oceg.org/critical-questions-about-ai/.

34. Pizzo. J. Thank you to Jay Scanlan, Priya Srinivasan, and the Accenture team for a valuable and interesting session!. LinkedIn. https://www.linkedin.com/posts/jessicapizzo_court-square-ai-teach-in-event-activity-7161067445619740674-BomX/?utm_source=share&utm_medium=member_desktop.

35. *Telus*. (n.d.). *We are committed to responsible AI*. https://www.telus.com/en/about/privacy/responsible-ai.

36. *Norges Bank Investment Management (NBIM)*. (2023). *Responsible artificial intelligence*. https://www.nbim.no/en/publications/our-views/2023/responsible-artificial-intelligence/.

37. The Model Mandate is broadly applicable and does not provide definitive guidance on individual issues – it can be repurposed to address responsible artificial intelligence (RAI).

38. Responsible Innovation Labs. (2023). *Voluntary Responsible AI Commitments for Startups & Investors*. https://assets-global.website-files.com/614b3d99f32e4ddc002ec49b/6553ef35b91b4d1d88c2c4d0_Responsible%20AI%20Commitments.pdf

39. The content of this case study was prepared solely for educational and informational purposes and does not constitute a recommendation, professional advice, an offer, solicitation or an invitation by or on behalf of Manulife IM or its affiliates to any person to buy or sell any security or strategies. Investment involves risks. In considering any investment, please consult a professional adviser. Investors for a sustainable digital economy and Manulife IM are not affiliated, and neither is responsible for the liabilities of the other.

40. *Creating Future Us*. (n.d.). *Sectors & Impacts of Focus*. https://creatingfutureus.org/technologies-sectors-of-focus/.

41. *Candriam*. (n.d.). *Insights: Facial Recognition*. https://www.candriam.com/en/professional/insight-overview/topics/facial-recognition/.

42. Chekroun, B., & S. Deleuze. (2022). *Investor Engagement on Human Rights Risks of Facial Recognition Technology: Interim Report*. Candriam. https://www.candriam.com/en/professional/insight-overview/topics/esg/facial-recognition-and-human-rights--responsible-investors-acting-together/.

43. *Norges Bank Investment Management (NBIM)*. (2023). *Responsible artificial intelligence*. https://www.nbim.no/en/publications/our-views/2023/responsible-artificial-intelligence/.

44. *Organisation for Economic Co-operation and Development (OECD)*. (n.d.). *Catalogue of Tools & Metrics for Trustworthy AI*. https://oecd.ai/en/catalogue/metrics.

45. *Stanford University*. (2024). *Holistic evaluation of language models (HELM).* https://nlp.stanford.edu/helm/instruction-pilot_six_scenarios/.

46. Velu, S., et al. (2024). *Adapting Your Organization for Responsible AI*. Bain & Company. https://www.bain.com/insights/adapting-your-organization-for-responsible-ai/.

47. Bevilacqua, M. N. Berente, H. Domin, B. Goehring & F. Rossi. (2023). *The Return on Investment in AI Ethics: A Holistic Framework*. arxiv.org. https://arxiv.org/pdf/2309.13057.

48. Hytha, M., & N. Kumar. (2022). *Activist Bluebell Urges BlackRock to Oust Fink as CEO Over ESG*. BNN Bloomberg. https://www.bnnbloomberg.ca/activist-bluebell-urges-blackrock-to-oust-fink-as-ceo-over-esg-1.1855751.

49. Green, J. & S. Foxman. (2024). *Why Corporate America Has a Diversity Problem*. Bloomberg UK. https://www.bloomberg.com/news/articles/2024-03-14/video-why-corporate-america-has-a-diversity-problem?srnd=homepage-canada.

The World Economic Forum,
committed to improving
the state of the world, is the
International Organization for
Public-Private Cooperation.

The Forum engages the
foremost political, business
and other leaders of society
to shape global, regional
and industry agendas.

World Economic Forum
91–93 route de la Capite
CH-1223 Cologny/Geneva
Switzerland

Tel.: +41 (0) 22 86 212
Fax: +41 (0) 22 7   2744
contact@wefor
www.weforum.