QUILTER CHEVIOT

RESPONSIBLE INVESTMENT

Moral machines: artificial intelligence, governance and ethics

Across Quilter we have identified three thematic engagement priorities. This is part of our human rights theme.

Human rights are rights inherent to all human beings, regardless of race, sex, nationality, ethnicity, language, religion, or any other status. Human rights include the right to life and liberty, freedom from slavery and torture, freedom of opinion and expression, the right to work and education, and many more. Everyone is entitled to these rights, without discrimination.

SDG Alignment

> *"AI is a powerful tool, but it's just a tool. It's up to us to decide how to use it."*
>
> Yoshua Bengio (Computer scientist and deep learning pioneer)

Introduction

Artificial intelligence (AI) is moving from experiment to infrastructure. It is being embedded in products, services, risk models, research tools, customer journeys, back-office systems and employee workflows. The question for investors is therefore no longer simply whether companies are "using AI", it is whether they understand where it sits, what decisions it shapes, what controls surround it, and who is accountable when it fails[1].

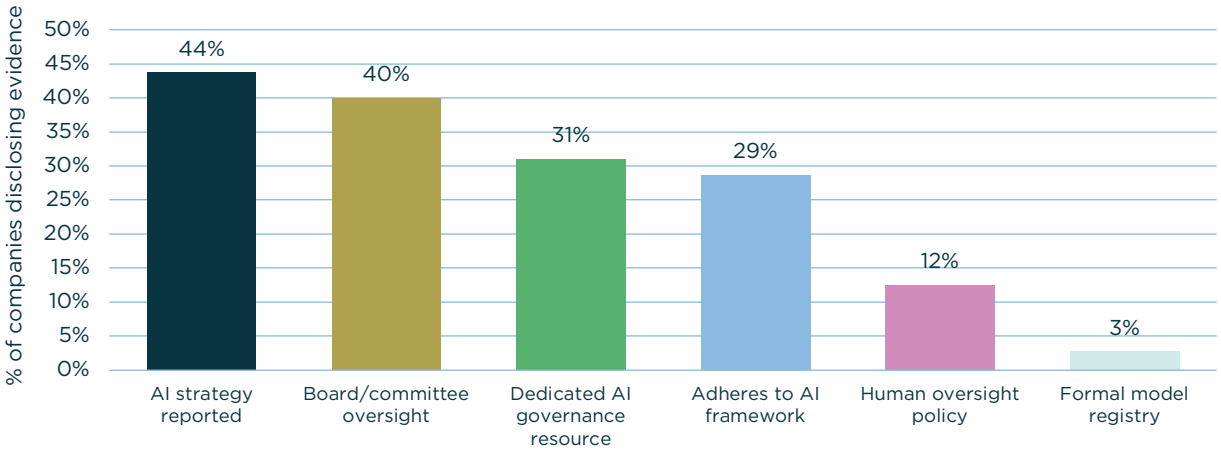
The governance gap is already visible. The AI Company Data Initiative (AICDI) is a third-party organisation gathering data on AI governance performance from almost 3,000 companies. The AICDI 2025 report found that 43.7% of companies publicly communicate having an AI strategy, but only 13% say they align that strategy with a formal AI governance framework; 40% report board or committee-level oversight, while only 12.4% report having a policy to ensure human oversight of AI systems and just 2.7%[2] publicly report a formal AI model registry[3]. It's clear that AI deployment is happening faster than corporate disclosures and governance frameworks can keep up with, and this matters because AI governance can affect factors from operational resilience to workforce stability and ultimately long-term company value.

1 AICDI, Responsible AI in practice: 2025 global insights from the AI Company Data Initiative, March 2026 (p.11)

2 AICDI, Responsible AI in practice: 2025 global insights from the AI Company Data Initiative, March 2026 (p.29)

3 An AI model registry is a centralised "system of record" that manages the complete lifecycle of machine learning models and AI assets. It acts as a single source of truth, tracking model versions, training data lineage, performance metrics, and deployment stages.

Responsible AI governance: the disclosure gap
(Selected AICDI 2025 indicators across 2,972 companies)



The risks are not theoretical. Biased algorithms, opaque decision-making and poor privacy controls can create legal, reputational and societal risks. In 2019, Apple's credit card, issued by Goldman Sachs, was investigated for offering significantly lower credit limits to women than men, despite almost identical financial profiles. In 2023, UnitedHealth faced a class action lawsuit alleging its AI tool systematically denied elderly patients access to extended care, with a human arbiter overturning the majority of cases on appeal. Privacy concerns are also mounting. Italy temporarily banned ChatGPT in 2023 over GDPR violations, citing unlawful data collection and lack of age controls. As AI systems increasingly rely on vast datasets, companies must ensure compliance with data protection laws and ethical data sourcing. Meanwhile, the EU's AI Act has begun enforcement, with fines of up to 7% of global turnover for non-compliance, signalling a new era of regulatory scrutiny. AI ethics has therefore become a financially material governance issue, shaping not only whether companies can innovate safely, but whether they can retain trust, meet regulatory expectations and keep human accountability intact as automation becomes more embedded in real-world decisions.

Evolutionary comparisons have been drawn to other major risks facing companies and societies, such as climate change. Climate disclosure eventually gained momentum through common language and regulatory backing as well as investor pressure and governance expectations, but AI lacks a universal metric like greenhouse gas emissions and has no governance framework equivalents like the Taskforce for Financial-related Climate Disclosures (TCFD)[4]. For AI, governance quality (i.e. board understanding, management accountability and systematic risk assessment) may be the most reliable leading indicator of responsible practice that we currently have.

Our engagement

We engaged AstraZeneca, Experian, HSBC, JPMorganChase, Merck, RELX and Wolters Kluwer to understand how AI ethics is being managed in practice across four areas:

Bias and fairness

Transparency and explainability

Privacy and data governance

Accountability and oversight

These companies were selected because they are significant AI users in high-impact sectors.

The companies we engaged with were not casual adopters. They are using AI in fundamental business processes in drug discovery, regulatory documentation, credit analytics, fraud detection, financial crime controls, customer service, investment research, clinical decision support, legal research and professional workflow tools.

Investors are increasingly clear that AI deployment is financially material. While attention often focuses on developers, most portfolio exposure sits with companies using AI across ordinary business processes, products and decisions. These deployer companies sit closest to clients and often face the largest liability risk related to any AI failures.

While the focus of our engagement looked at the four key practice areas mentioned above, AI ethics and governance risks also cut across familiar ESG themes, including nature, climate and human rights. On social issues, AI is already reshaping work. Some companies have cut roles on the assumption that AI will deliver efficiencies, only to rehire when

4 Great Circle Capital Advisors and Omidyar Network, 'From carbon to code: applying lessons from a decade of climate finance practice to responsible investor engagement on AI', March 2026

those gains fail to materialise (the so-called 'AI Boomerang' effect). There is also a wider question about whether entry-level roles are being eroded, and what that means for the next generation of workers.

As Quilter Cheviot explores in our 'Greening Algorithms' thematic engagement, environmental risks are also rising. AI increases demand for water, land, energy and associated emissions. This is not only a hyperscaler issue: companies that use AI can bring these impacts into their own supply chains, creating new exposure for sectors such as finance that have traditionally viewed emissions mainly through an investment lens.

Encouragingly, throughout our engagements, all companies framed AI as something requiring governance, not just enthusiasm. The stronger examples treated AI as an enterprise risk and trust issue rather than a technology project.

Is AI ethical?

This is a question that is more regularly appearing on the horizon, so worth discussing. AI is not ethical or unethical in the abstract. It is a tool - or increasingly, a system of tools - that becomes more or less ethical depending on where it is used, what it is trained on, who controls it, what incentives surround it, and whether affected people can understand, challenge or remedy outcomes. A machine learning model that helps identify drug targets under expert supervision is not the same ethical proposition as an opaque tool affecting credit access, healthcare coverage or employment outcomes without explanation or appeal.

The ethical challenge is that AI can scale both good and bad decisions. RELX's ClinicalPath product, for example, supports oncologists by citing clinical guidelines and keeping final approval with doctors; Wolters Kluwer's VitalLaw and UpToDate Expert AI tools ground outputs in curated sources and human validation; AstraZeneca and Merck emphasise human-centric deployment in high-stakes pharmaceutical and patient-facing settings. These are examples of AI being used as an expert-support tool, not a substitute for judgement.

But without governance, AI can also scale bias, privacy failures and unaccountable decisions. As mentioned above, there is limited publicly available evidence of model registries, impact assessments, human oversight policies and external AI ethics advisory boards across the wider market. The ethical verdict therefore depends less on the model than the management system. AI can be ethical only where governance makes it so.

AI can be ethical only where governance makes it so.

AI governance risks do not exist in isolation. Weak oversight of AI systems can amplify broader environmental and social risks emerging across the value chain, particularly as demand for data centres accelerates. Collectively, data centres already consume more electricity than all but the world's largest countries and are projected to see demand rise sharply as AI adoption grows. This expansion is increasingly encountering local resistance, driven by concerns over energy use, water consumption, land requirements and the construction of dedicated power infrastructure. In the US alone, 75 data centre projects worth approximately $130bn were blocked or delayed in the first quarter of 2026[5]. These tensions highlight a wider trust challenge. Public concerns about algorithmic bias, privacy and accountability are increasingly intersecting with concerns about the physical footprint of AI infrastructure. Companies that fail to provide transparency around both AI governance and the environmental and social impacts of deployment risk regulatory scrutiny, project delays and damage to their social licence to operate. Effective governance is therefore not only about managing AI systems responsibly but also maintaining public trust in the wider ecosystem that supports them[6].

Regulatory developments: governance moves centre stage

The regulatory landscape for AI is evolving rapidly, with the EU AI Act emerging as the most significant global framework and the US taking a more fragmented, sector-led approach. The EU AI Act is the world's first comprehensive AI regulation. It takes a risk-based approach, categorising AI systems according to their potential impact on individuals and society.

Unacceptable-risk systems (such as certain forms of social scoring) are prohibited. High-risk systems used in areas including healthcare, employment, education, critical infrastructure and financial services face extensive requirements. These requirements for high-risk applications include risk assessments, human oversight, data governance controls, transparency requirements, technical documentation, incident monitoring and accountability mechanisms.

Non-compliance can result in fines of up to 7% of global annual turnover for the most serious breaches. Its influence

5 The Great AI Data Centre Cover-Up, Financial Times, 8 July 2026

6 PRI, Managing AI Governance Risk: A Practical Guide for Investors, 2026

extends well beyond Europe. AICDI found that among companies publicly disclosing adherence to AI frameworks, 53% reference the EU AI Act, making it the dominant governance benchmark globally.

Unlike Europe, the US has no single federal AI law. Instead, regulation is developing through a combination of sectoral regulation, state legislation and existing legal frameworks.

Key developments include:

- The National Institute of Standards and Technology (NIST) AI Risk Management Framework (AI RMF), which has become a widely used voluntary governance framework for identifying, assessing and managing AI risks.
- The Colorado AI Act, which introduces obligations for developers and deployers of high-risk AI systems, including documentation and discrimination-risk mitigation requirements.

For many companies, the challenge is no longer anticipating regulation but preparing for a patchwork of overlapping requirements.

How companies are responding to regulatory developments

Our engagement suggests leading companies are already preparing for this environment. Several firms referenced the EU AI Act when discussing governance structures, risk assessments and deployment controls.

Healthcare companies and those providing research services appear particularly focused on explainability and human oversight. Wolters Kluwer highlighted that clinical AI applications such as 'UpToDate Expert AI' are designed around expert-reviewed content, source attribution and clinician oversight rather than fully autonomous decision-making. The company also noted the growing availability of open-source medical AI applications, which may accelerate innovation but often operate outside established governance, validation and accountability frameworks. Some of these applications are also advertising platforms for pharmaceutical products, adding to the risk of ethical failings. This reinforces the value of trusted governance structures using high quality data in high-stakes sectors such as healthcare.

More broadly, companies are increasingly embedding AI governance within existing risk, compliance, privacy and model-risk frameworks rather than treating AI ethics as a standalone initiative.

Regulation is moving governance from a voluntary best practice to a business necessity. Companies that already have robust oversight, impact assessments, human accountability and transparent controls are likely to be better positioned as regulatory expectations continue to increase.

What is responsible AI?

There is no single agreed metric for responsible AI. As referred to above, other material systemic risks like climate change benefitted from a relatively clear common metric - emissions - while AI risks are multidimensional, use-case specific and fast evolving[7]. Even on broader measures, companies often describe governance at a conceptual level but provide less evidence of how controls function across the lifecycle of AI systems[8], so benchmarking can be difficult.

Aligned with best practice frameworks, such as the EU AI Act and The Organisation for Economic Co-operation and Development (OECD) AI Principles, for our engagement, responsible AI principles mean AI that is purposeful, fair, explainable where needed, privacy-protective, monitored, contestable and accountable to people. In practice, this requires four pillars:

- **Bias and fairness**: testing datasets and outputs for discriminatory or unfair outcomes. Bias is most material where AI affects access to services, healthcare, credit, employment, security or legal outcomes.
  - ***Examples***: AstraZeneca tests for unfair bias in datasets and recognises the need to improve diversity in areas such as genomic research. Experian integrates fairness testing, explainability assessments and model performance reviews throughout the AI lifecycle, which is particularly important given its role in credit, fraud and identity-related decisions. JPMorganChase's fairness work sits within a broader model-risk and control framework, supported by the Trustworthy AI Center of Excellence, but external disclosure of bias-testing outcomes remains limited.
- **Transparency and explainability**: ensuring users, experts, auditors and regulators can understand or interrogate AI-assisted outputs. Explainability is not just a technical preference; it is a trust mechanism. The common gap is

7 Great Circle Capital Advisors and Omidyar Network, 'From carbon to code: applying lessons from a decade of climate finance practice to responsible investor engagement on AI', March 2026

8 AICDI, Responsible AI in practice: 2025 global insights from the AI Company Data Initiative, March 2026 (p.29)

external reporting: companies can often explain outputs internally or to clients but provide investors with limited aggregated evidence of model performance, limitations or error rates.
  - ***Examples***: RELX's generative AI legal products use curated databases and source references to improve traceability. Wolters Kluwer's VitalLaw provides citations and human editorial checks, while UpToDate Expert AI gives evidence pathways for clinicians.

- **Privacy and data governance**: controlling data origins, access, consent, security, retention, vendor sharing and model training. Privacy has a head start because it sits on mature legal and compliance infrastructure. Banks, credit analytics firms, pharmaceuticals and professional information businesses already operate with strict data obligations. The next test is not whether privacy policies exist but whether they cover model training, third-party data flows, prompt inputs, output reuse, deletion rights and vendor auditability.
  - ***Examples***: HSBC uses sandbox environments to reduce sensitive data leakage and controls interactions with external models. JPMorganChase operates 'LLM Suite', an internal AI platform, in a controlled environment and indicated that client data is not used to train models.
- **Accountability and oversight**: ensuring humans remain responsible, high-risk systems are reviewed, and incidents can be escalated, remediated or withdrawn. Accountability is where responsible AI either becomes real or collapses into branding. Are there designated bodies reviewing accountability? Are there dedicated resources with authority, lifecycle traceability, user feedback mechanisms and impact assessments? Our engagement evidence suggests leading companies are building these systems, but public comparability remains poor.

What best practice looks like today

Bias & fairness	Transparency & explainability	Privacy & data governance	Accountability & oversight

How to read: the circular marker is a four-part topic ring. Coloured segments and row tints show the themes each best-practice area most clearly addresses.

	Best-practice area	Topic alignment	Investor signal	Engagement evidence
01	Board and executive oversight	Accountability & oversight	Regular board or committee attention, senior ownership, AI-literate directors or access to expert advice.	AstraZeneca has board and executive oversight; Experian's Global AI Risk Council escalates to senior committees and the board where appropriate; HSBC has a Chief AI Officer and risk committee integration.
02	Formal governance framework	Accountability & oversight Transparency & explainability	Clear AI policy or principles aligned with recognised frameworks such as OECD, UNESCO, NIST or the EU AI Act.	Merck's principles-based policy is informed by OECD AI Principles and the EU AI Act; RELX and Wolters Kluwer have Responsible AI Principles; Experian has a Global AI Policy.
03	AI inventory / model registry	Privacy & data governance Accountability & oversight	A central record of AI systems, owners, risk classification, data sources, vendors and monitoring status.	AI model registries are rarely evidenced across the market, making this a clear engagement priority.
04	Impact assessments	Bias & fairness Privacy & data governance Accountability & oversight	Ethical, human rights, privacy, data protection and environmental impact assessments, repeated when use cases change.	Merck requires every AI system to undergo an AI impact assessment, broader than traditional risk assessment and repeated when systems are updated or implemented elsewhere.
05	Human oversight	Accountability & oversight	Named human accountability, review rights, override routes, pause/rollback authority and no blame deferred to algorithms.	AstraZeneca, Merck, HSBC, RELX, Wolters Kluwer and JPMorgan Chase all described human oversight in high-impact use cases.
06	Explainability and contestability	Transparency & explainability Accountability & oversight	Source citations, reason codes, audit trails, user challenge routes and disclosure where AI is used.	RELX and Wolters Kluwer use curated sources and evidence pathways; Experian states impacted individuals should have routes to question or challenge AI-assisted outcomes.

	Best-practice area	Topic alignment	Investor signal	Engagement evidence
07	Data and vendor controls	Privacy & data governance Bias & fairness	Dataset origin tracing, bias testing, permissioning, vendor due diligence, audit rights and secure AI sandboxes.	AstraZeneca uses guardrails and partner due diligence; HSBC uses a model-agnostic gateway and sandboxing; JPMorgan Chase uses controlled deployment and role-based access.
08	Incident escalation and remediation	Accountability & oversight	Clear reporting, escalation to senior leadership, incident registers and remediation pathways.	AstraZeneca and RELX described escalation routes for AI-related incidents.
09	External assurance and transparency	Transparency & explainability Accountability & oversight	Independent audits, external advisory panels, public case studies, selected metrics and disclosure of governance outcomes.	Merck's Digital Ethics Advisory Panel is a positive feature; AICDI finds external AI ethics advisory boards and third-party bias assurance remain rare across the wider market.
10	Workforce preparedness	Bias & fairness Accountability & oversight	AI literacy, role-specific training, worker safeguards and grievance mechanisms for AI-related workplace impacts.	AICDI found only 31% of companies evidence AI training/ reskilling and much less evidence of comprehensive coverage or grievance mechanisms, making workforce governance a clear area for future engagement.

Note: Best practice areas and signals outlined by AICDI

Key findings from company engagement

1 Governance is the leading indicator

For AI, governance quality may be more useful than fixed metrics because the technology and risk landscape changes too quickly for static targets to capture what matters[9]. This was visible in our engagements. Merck requires every AI system to undergo an AI impact assessment broader than a traditional risk assessment, with each principle broken into actionable questions and reassessment when systems are updated or deployed in new areas. AstraZeneca has implemented AI governance and AI risk management frameworks, with board and executive oversight, high-impact system controls and post-deployment monitoring.

The best approaches were not isolated in technical teams. Experian's Global AI Risk Council reports into senior leadership, the Executive Risk Management Committee and, where appropriate, the Board and Audit Committee. HSBC integrates AI oversight into existing risk structures, has a cross-functional AI Review Committee, a Chief AI Officer, an AI Centre of Excellence and applies its Three Lines of Defence model to AI. JPMorganChase embeds AI within data governance, privacy, cybersecurity, model-risk management and operational resilience rather than treating it as a standalone ethics line.

2 Human oversight is the safety rail

Human oversight appeared repeatedly as the practical bridge between innovation and accountability. AstraZeneca views AI as augmenting scientists and decision-makers, with human review retained for important decisions in drug development, patient care and even highly automated supply chain scenarios. Merck uses a human-AI partnering approach and keeps high-stakes, real-time patient-facing applications under human oversight. HSBC explicitly avoids fully autonomous AI in critical processes, reflecting regulatory expectations and internal risk appetite.

RELX and Wolters Kluwer offer particularly clear product-level examples. RELX's ClinicalPath product provides AI-supported cancer treatment guidance but requires doctor approval, while Wolters Kluwer's human/expert-in-the-loop approach ensures high-stakes outputs remain subject to qualified professional judgement. JPMorganChase described proxy voting as another constructive model: AI supports research, but final voting decisions remain with portfolio managers and stewardship teams.

9 Great Circle Capital Advisors and Omidyar Network, 'From carbon to code: applying lessons from a decade of climate finance practice to responsible investor engagement on AI', March 2026

3 Strong performance in data governance

Privacy and data governance were among the strongest areas across engagements, especially among companies handling regulated or sensitive data. Experian's AI governance is built on longstanding Global Data Principles, privacy-by-design practices, access controls, audit trails, consumer dispute mechanisms and data quality oversight. AstraZeneca relies primarily on internal and partner datasets, restricts confidential data use in external AI tools and uses guardrails and secure sandboxes for generative AI experimentation. Wolters Kluwer grounds its 'Expert AI' approach in proprietary, expert-curated data and human validation, and avoids indiscriminate scraping of personal data from open sources.

This matters because AICDI finds that training-data and third-party data controls are weak across the wider market. It reports that only a minority of companies have policies for training-data quality, third-party AI data sharing or user rights. It warns that without contractual limits on reuse and audit rights - biased or non-compliant data can remain opaque to deployers using third-party models.[10]

4 Transparency is the unfinished work

Most companies we engaged appeared stronger on internal governance than external disclosure. Experian discloses policies and oversight structures but less on bias testing, incidents, appeals, model performance and post-deployment monitoring. JPMorganChase has significant technical capability and a Trustworthy AI Center of Excellence, but we did not find consolidated public reporting on production-level fairness testing, model performance or hallucination controls. Wolters Kluwer provides strong explainability within products, but does not yet publish external model accuracy, error-rate or AI ethics case-study reporting.

Investors need evidence of how models are approved, monitored, corrected and, where necessary, withdrawn, not simply statements of intent. This is the next disclosure frontier. We should not expect every company to publish commercially sensitive model details, but we should expect more outcome-based reporting, selected case studies, incident processes and governance metrics.

5 Best practice is emerging - but unevenly

The most advanced approaches combine formal principles, board oversight, impact assessments, human accountability, curated data, vendor controls, lifecycle monitoring and clear escalation. Merck's external Digital Ethics Advisory Panel is a notable governance feature, providing independent advice on complex ethical challenges. RELX uses Responsible AI Principles, cross-functional senior management oversight and product-level human accountability but could still improve public transparency on audits and metrics. Wolters Kluwer's central Digital eXperience Group, FAB platform (internal tool to help scale AI processes), Responsible AI Principles and AI Centre of Excellence provide a strong enterprise governance model, though more formal board-level AI ethics oversight and external reporting would strengthen accountability.

Who appears ahead - and where gaps remain

Pharmaceuticals stood out for caution and process discipline. AstraZeneca and Merck both emphasised human-centric AI, cautious deployment in high-stakes patient-facing contexts, strong privacy controls and structured governance. Merck's universal AI impact assessment and external Digital Ethics Advisory Panel are particularly close to current best practice. This aligns with the sector position as a high-risk space for AI deployment under the EU AI Act.

Financial services showed mature risk architecture, but weaker public transparency. HSBC, JPMorganChase and Experian benefit from deep model-risk, privacy, compliance and cybersecurity cultures. HSBC's Three Lines of Defence model and Chief AI Officer, JPMorganChase's Trustworthy AI Center of Excellence and Experian's Global AI Risk Council are credible structures. The gap is less about whether controls exist and more about whether investors can see enough evidence of outcomes, testing, incidents, model performance and human-review processes.

Professional information companies showed strong product-level explainability. RELX and Wolters Kluwer have a clear understanding that trust is core to their business model. Curated data, traceable outputs, human validation and expert-in-the-loop design are strong features. The next step is fuller public reporting on AI accuracy, limitations, audit findings and governance metrics. There appears to be an inherent governance advantage in having close control of proprietary, curated datasets which feed AI based products, rather than the higher risk outputs derived from general large language models.

10 AICDI, Responsible AI in practice: 2025 global insights from the AI Company Data Initiative, March 2026

Across all sectors, the recurring weaknesses are similar: outcome-based reporting, impact assessments, formal model inventories, incident transparency and external assurance. These are not reasons to reject AI, they are reasons for sustained stewardship.

Summary

We were encouraged by the seriousness of the companies engaged. The overall picture is not one of reckless deployment, but nor is it one of mature market-wide assurance. The better companies are moving in the right direction by embedding AI into enterprise risk management. But even among sophisticated users, disclosure often lags practice, and investor visibility into outcomes remains limited.

This is the AI ethics equivalent of the early climate disclosure era. The systems are being deployed, the risks are material, the frameworks are proliferating, and the metrics remain imperfect. Early path dependencies matter and the responsible AI field should therefore avoid rushing into weak metrics while underinvesting in governance, data infrastructure and implementation support[11].

The central lesson is that responsible AI cannot be reduced to a policy document. It depends on a control environment that works in practice - one where Boards ask informed questions and management owns the risks. Technical teams will document systems properly, while legal and compliance functions track regulation. Importantly, employees will understand how to use AI safely and users can receive meaningful explanations. Ultimately, human accountability should remain clear whenever automated systems shape real-world outcomes.

Practically, we will continue to engage companies on this topic and use these findings to set more formal expectations. One outcome will certainly be a clearer voting policy alignment in supporting well-written, meaningful shareholder resolutions aiming to improve AI governance performance and disclosure.

AI's potential is vast. So are the risks. The pathway to a positive outcome is not blind acceleration or blanket scepticism, but disciplined governance. If companies can pair innovation with accountability, AI can become not just a productivity engine, but a trust-building technology. If they cannot, the same tools may amplify bias, opacity and fragility at scale. As with climate, the winners will not simply be those with the boldest ambitions, but those with the governance to deliver them.

11 Great Circle Capital Advisors and Omidyar Network, 'From carbon to code: applying lessons from a decade of climate finance practice to responsible investor engagement on AI', March 2026

Greg Kearney
Senior Responsible Investment Analyst

Quilter Cheviot
Senator House
85 Queen Victoria Street
London EC4V 4AB

+44 (0)207 150 4000 **enquiries@quiltercheviot.com** **quiltercheviot.com**