Amundi
Investment Solutions

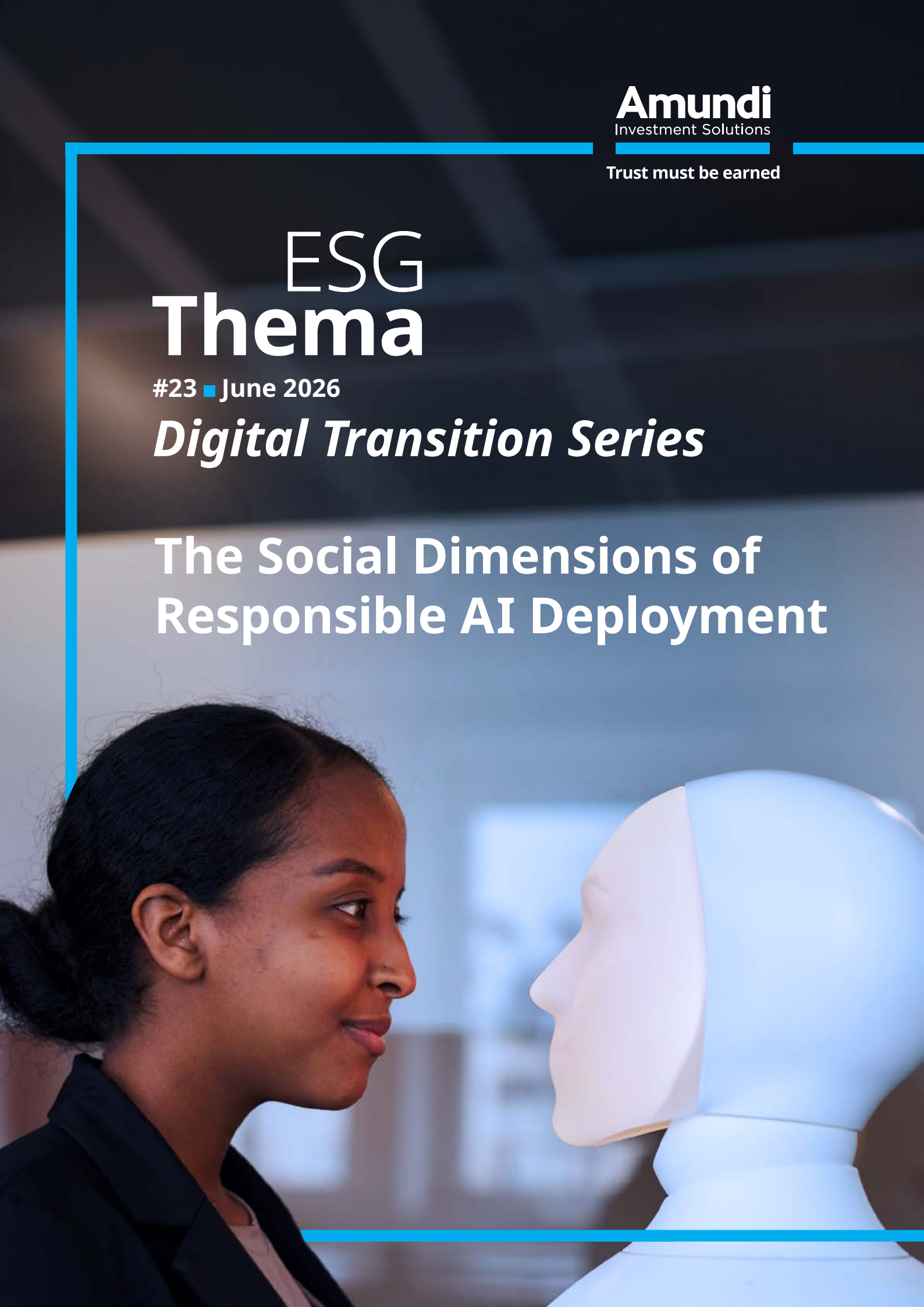
Trust must be earned

ESG
Thema

#23 ■ June 2026

Digital Transition Series

The Social Dimensions of Responsible AI Deployment



Key takeaways

- **AI is a material source of economic opportunity, but its benefits will depend on responsible deployment:** AI can support productivity, innovation and efficiency, yet weak governance of AI-related risks may undermine these gains. This leads to a strong case for investor engagement.
- **A governance gap continues to widen between AI commitments and implementation:** Many companies now have ethical AI principles, but disclosure on practical controls remains limited and lags considerably behind AI diffusion.
- **Social risks arise across the AI value chain, with deployment emerging as a key investor focus:** Developers shape model behaviour, but a large share of investors' portfolio-level social risk sits with companies deploying AI across established workplace processes, products, services and consequential decisions.
- **AI is reshaping work in ways that extend beyond headcount effects:** The main risks relate not only to displacement, but also to work redesign, weaker worker voice, algorithmic management and the erosion of formative early career tasks that build future human judgement and oversight capacity.
- **Downstream harms are becoming more visible and more material:** AI-related discrimination in access to services, child safety failures, dual-use applications and community tensions linked to data centre expansion show how weak controls can create harm at both individual and systemic levels.
- **Investor stewardship should focus on evidence of a governed AI lifecycle:** Investors should expect companies to demonstrate meaningful human oversight of AI, with ongoing monitoring and viable routes to remedy and adaptation as AI evolves.

Introduction

AI has become a material source of economic opportunity and innovation.[1] OECD economists argue that it can raise productivity by improving business processes, output quality, innovation and the efficiency with which firms use labour and capital, under the right organisational and policy conditions.[2] However, AI diffusion also introduces risks that can undermine its potential for positive economic and wider societal impact if not managed through an ethical approach to development and deployment. Investors therefore need to not only assess these risks and monitor their evolution, but also act upon them through meaningful and purposeful engagement.

Recent analysis highlights a growing responsible AI governance gap: the distance between public commitments to ethical AI and evidence that those commitments are embedded in governance and practice. This concern has become more material as AI adoption and capability have accelerated.[3] Indeed, corporate AI governance disclosure suggests that implementation is lagging stated intent. The latest report produced by the Thomson Reuters Foundation in partnership with UNESCO finds that companies are adopting AI faster than their governance frameworks are adapting: nearly 90% of companies studied have not publicly committed to a named AI governance framework, only 13% disclose a policy to ensure human oversight of AI systems, and just 2.3% have a dedicated complaints mechanism for AI-related issues.[4] World Benchmarking Alliance research on 200 major technology companies reaches a similar conclusion: 38% had public AI principles, 24 companies explained internal AI governance mechanisms, and no company showed proof of conducting comprehensive human rights impact assessments on the AI systems they developed, procured or deployed.[5]

The policy-to-practice gap is also becoming harder to close because AI systems are changing quickly and remain difficult to assess. The 2026 International AI Safety Report notes that performance in pre-deployment tests does not reliably predict real-world AI utility or risk, internal models remain poorly understood, and developers have incentives to keep important information proprietary.[6] OECD analysis adds a lifecycle concern, whereby updates to general-purpose AI models can significantly alter capabilities and risk profiles, yet often avoid comprehensive assessment, even when downstream users already depend on those models. Its analysis of data from 143 major providers found that over 60% of updates could potentially increase systemic risk, while just under 5% focused on improving safety and security mitigations.[7]

Together, the governance gap and the opacity of AI systems themselves present a compounding stewardship challenge for investors, one that requires tools for identifying where social risks are concentrated across the AI value chain and how the agenda should adapt as AI capability continues to develop. In this paper, we examine the social dimensions of AI impact through a value chain lens, applying it to several dimensions that have grown in materiality. We discuss the tools and frameworks that companies, both AI developers and deployers, have at their disposal to mitigate the relevant risks, how scenario analysis can inform stewardship as the AI risk landscape evolves, and the engagement expectations that investors can bring to their dialogue with companies.

1. See research by Amundi Institute on the diverse economic impacts of artificial intelligence. Available at https://research-center.amundi.com/article/diverse-economic-impacts-artificial-intelligence
2. Filippucci et al., 2025. "Opportunities and Risks of Artificial Intelligence for Productivity." International Productivity Monitor, Number 48, Spring 2025. OECD Economics Department. Available at https://www.csls.ca/ipm/48/OECD_Final.pdf
3. Stanford HAI. 2026. AI Index Report 2026. Stanford Institute for Human-Centered Artificial Intelligence. Available at: https://hai.stanford.edu/ai-index/2026-ai-index-report
4. AI Company Data Initiative (AICDI) / Thomson Reuters Foundation. 2025. Corporate AI Governance Report 2025. Available at https://www.trust.org/resource/ai-company-data-initiative-2025-insights/. See also: AICDI. 2026. Responsible AI in Practice. Available at https://www.trust.org/wp-content/uploads/2026/03/AICDI-2025-Responsible-AI-in-practice-1.pdf
5. World Benchmarking Alliance. 2026. "Tech sector progress on AI accountability threatens to stall." Available at https://www.worldbenchmarkingalliance.org/tech-sector-progress-ai-accountability-threatens-stall
6. International AI Safety Report. 2026. International AI Safety Report panel / AI Safety Institute. https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026
7. IOECD.AI. 2025. "Proportional oversight for AI model updates can boost AI adoption." OECD AI Policy Observatory. Available at https://oecd.ai/en/wonk/proportional-oversight-for-ai-model-updates-can-boost-ai-adoption.

Figure 1. A framework for investor stewardship on the social dimensions of AI

AI presents material economic opportunity and social risk. Responsible AI development and deployment requires active investor stewardship:

Apply AI scenario analysis

Assess material risks	Understand governance levers and responsibility	Engage with companies
Risk categories along the value chain: Upstream Developer Deployer Downstream Societal	**Developers:** Model design Training data governance Documentation and responsible use policies **Deployers:** Vendor due diligence Deployment decisions Use case governance Human oversight Remedy architecture **+ Shared developer-deployer responsibility**	**Engagement expectations for developers and deployers:** Know your AI Control your AI Account for AI impacts Govern your AI

I. The AI social value chain

AI development and deployment involve a complex chain of activities, each of which generates distinct social risks, which investors in the AI systems involved need to consider. Developers shape model behaviour, training data governance and documentation quality, whereas deployers decide where and how AI is used and how much meaningful control people retain in practice. To date, the largest public attention has fallen on frontier AI developers, yet the larger share of portfolio-level social risk and opportunity exposure sits with companies deploying AI into ordinary business processes. The figure below identifies where risks arise in the chain social, and which actor bears primary responsibility for managing them.

Figure 2. The AI value chain: where social risks arise

STAGE	KEY PROCESSES	PRIMARY ACTOR	MAIN SOCIAL ISSUES AND INVESTOR RELEVANCE
Hardware inputs	Extraction and processing of material inputs for AI hardware and supporting infrastructure	Primarily developers and infrastructure operators	Labour and community harms in high-risk sourcing contexts.
Component manufacturing and assembly	Manufacturing and assembly of chips, servers and related equipment	Primarily developers and infrastructure operators	Supply chain opacity can obscure factory working conditions. Weak component-level safeguards can facilitate downstream product misuse.
Data governance, annotation and moderation	Collection, selection and governance of training data; data labelling, content review and human feedback	Primarily developers	Privacy, bias and data legitimacy enter systems before deployment. Data annotation and moderation work concentrates psychological and labour risks, often in opaque low-income markets.
Model development and evaluation	Model training, testing, documentation and release	Primarily developers	Safety, misuse and dual-use risks are shaped at this stage. Weak documentation and narrow evaluation scope limit deployer and investors risk assessment. Governance issues here cascade downstream.
Deployment and enterprise integration	Use of AI in products, services and internal decision-making	Primarily deployers	The stage at which AI enters ordinary work, services and consequential decisions. Work redesign, algorithmic management, oversight failures and consumer harm risks are concentrated here.
Infrastructure operation	Operation and expansion of data centre and cloud infrastructure	Both developers & deployers	Compute expansion creates concentrated resource demands on local resources and requires public consent that cannot be assumed.

Source: Amundi analysis.

The stages are interconnected, which is important for investor assessment. How developers design and document models determines what deployers can do with them and what human oversight remains feasible. Meanwhile, how companies deploy AI shapes both individual outcomes and wider labour market and rights effects. The consequences of this value chain for users, affected stakeholders and the wider society are examined in the sections that follow.

The themes selected for our analysis share three characteristics. They experience increasing regulatory and stakeholder scrutiny. All also sit in territory where existing investor frameworks leave significant analytical gaps, either because the risk is genuinely novel or because AI changes its character in ways that standard sector analysis does not reach. Finally, each of these risks compounds with the others: for instance, formative work erosion weakens the organisational capacity needed to oversee consumer-facing AI; governance failures in consequential decisions generate the population-scale harms that erode social licence; weak remedy architecture denies both workers and affected users practical routes to correction and restitution.

The interaction between these themes underscores the need for holistic investor stewardship on responsible AI. As well, the risks identified in the table above extend beyond the organisations that generate them. The cases examined in the following sections illustrate how inadequate controls at the deployment stage translate into harm at the level of individual users and affected communities. At greater scale, firm-level failures aggregate into systemic effects, and such effects exceed the sum of individual firm failures, which is part of what makes them a distinctive investor concern. They tend to build gradually and become visible only once the scale of harm has already made remediation difficult.

Examining every social risk in the AI value chain would be beyond the scope of this paper, and two prominent categories have been deliberately set aside. First, human rights risks associated with hardware inputs, including labour conditions in minerals extraction and component manufacturing, are already subject to established responsible business conduct frameworks and active investor engagement through existing supply chain due diligence mechanisms. Meanwhile, systemic risks of AI to democratic institutions, information integrity and geopolitical stability are acknowledged as material; this paper focuses on corporate-level social governance.[8]

II. Social risks inside firms and across the value chain

The following sections examine the social risks that are most material to investors across the AI value chain and are organised in three groups:

1. **Upstream risks arising in the development and supply of AI systems**
2. **Organisational risks arising inside firms as AI is deployed into work processes and decisions**
3. **Downstream risks that manifest for users, affected communities and the wider society**

1. Upstream issues

Supply chain and data labour

AI data labour refers to the work of annotating, labelling and categorising the data on which AI models are trained.[9] It is often performed by workers in low-income countries, often in precarious conditions.[10] This geographic distribution reflects a cost-sensitive business model that draws data work to jurisdictions with labour costs are lower and regulation tends to be less protective.[11] Reported pay for some AI data-

8. For discussion of these risks, see e.g., International AI Safety Report. 2026.
9. Gonzalez-Cabello et al. 2024. Fairness in crowdwork: Making the human AI supply chain more humane. Business Horizons. Available at: https://www.anderson.ucla.edu/sites/default/files/document/2025-05/Fairness%20ins%20crowdwork.pdf
10. GIZ / BMZ Digital. 2026. Invisible Workers, Visible Harms: Perils and Precarities of AI Labour. Bonn: Deutsche Gesellschaft fur Internationale Zusammenarbeit. Available at: https://www.bmz-digital.global/wp-content/uploads/2026/02/GIZ_2026_InvisibleWorkersVisibleHarms-1.pdf
11. Brookings Institution. 2025. "Reimagining the Future of Data and AI Labor in the Global South." 7 October 2025. Fairwork. 2025. Cloudwork Report 2025: Advancing Standards in Digital Labour Platforms. Oxford Internet Institute / Fairwork. Available at: https://fair.work/wp-content/uploads/sites/17/2025/05/Fairwork-Cloudwork-Report-2025-FINAL.pdf

labelling and moderation work in lower-cost markets has fallen below US$2 per hour, illustrating the strength of labour cost arbitrage in the sector.[12] Moreover, the sector is organised in ways that weaken accountability for working conditions. A common model uses layered outsourcing: large AI developers contract with business process outsourcing intermediaries, which may in turn rely on platform-based or subcontracted labour. This creates legal and operational distance between the companies that benefit economically from the work and the workers who perform it, with workers often classified as independent contractors and therefore outside the scope of standard employment protections such as minimum wage coverage, sick pay, unemployment insurance and collective representation.[13] Conditions experienced by AI data workers have been documented in a growing body of reporting and investigation.[14]

Workers have been organising in response, signalling an emerging labour movement in the AI supply chain. This is likely to bring growing regulatory, legal and reputational pressure to bear on AI developers and their institutional investors. The EU AI Act increases pressure on data governance and documentation,[15] while the OECD Due Diligence Guidance for Responsible AI more directly extends responsible business conduct expectations across AI supply chains.[16] **Investors should therefore seek to understand whether AI developers can demonstrate meaningful visibility over the labour conditions of data annotation and content moderation workers across layered supply chains, and whether contractual standards extend to subcontracted and platform-based workers rather than ending at the direct supplier.**

Content moderation and AI data labour

AI data labellers and content moderators are often discussed together but perform distinct work. Data labellers annotate the training data that makes AI models learn; content moderators review user-generated content on platforms to enforce community standards. The two intersect because moderators' decisions are frequently used to train AI safety classifiers, placing content moderation within the AI supply chain. Investor expectations are consistent across both: visibility over working conditions, contractual accountability and access to remedy, directed at the relevant actor: model developers for data labelling and platforms for content moderation.

2. Organisational issues

AI and the evolving world of work

Early discussions around AI and jobs was dominated by exposure estimates: how much of an occupation's or a role's task bundle could, in principle, be performed faster or more cheaply by AI? Recent analysis distinguishes between task exposure and job loss, suggesting that exposure can be misread as a forecast of job displacement.[17] Many jobs consist of complementary tasks, and automating some can raise the value of the remaining human work. AI exposure reflects potential changes to those work tasks, and the longer-term outcome for employment depends on how firms and workers respond. Exposure measures alone can obscure both workforce resilience and vulnerability. Manning and Aguirre's 2026 analysis finds that 37.1 million US workers are in the top quartile of AI exposure, but around 70% are in occupations with above-median adaptive capacity.[18] Workforce outcomes of AI diffusion will likely depend on the structure of individual roles and the scope workers have to adjust. Some roles may be reshaped or augmented, while leaving individual judgement and accountability with the workers affected. Automation and augmentation can appear within the same role, with some skills losing value and others gaining importance.[19]

12. Techglobal Institute. 2025. "Labor in the shadows: Rights and Risks for Asia's data workers". Available at: https://www.bmz-digital.global/wp-content/uploads/2026/02/GIZ_2026_InvisibleWorkersVisibleHarms-1.pdf
13. See GIZ, Brookings Institution, Fairwork, op cit.
14. Brookings Institution. 2025. Op cit. "Reimagining the Future of Data and AI Labor in the Global South." 7 October 2025
15. European Union. 2024. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence, Artificial Intelligence Act.
16. OECD. 2026. OECD Due Diligence Guidance for Responsible AI. OECD Publishing. Available at: https://www.oecd.org/en/publications/2026/02/oecd-due-diligence-guidance-for-responsible-ai_7831bb49.html
17. Gans, Joshua S., and Avi Goldfarb. 2026. "O-Ring Automation." NBER Working Paper No. 34639. Available at: https://www.nber.org/papers/w34639 Manning, S. & Aguirre, T. 2026. "How Adaptable Are American Workers to AI-Induced Job Displacement?" NBER working paper / NBER volume chapter. Available at: https://www.nber.org/papers/w34705
18. Manning, S. & Aguirre, T. 2026. Op cit.
19. Burning Glass Institute. 2026. Beyond the Binary: How Automation and Augmentation Are Shaping the AI Labour Market. 28 January 2026. Available at: https://www.burningglassinstitute.org/research/beyondthebinary

Institutional bodies are also moving towards a more nuanced position on workforce impacts of AI. The ILO's 2025 update states that one in four workers globally are in occupations with some degree of generative AI exposure, but that most jobs are more likely to be transformed than eliminated because human input remains necessary.[20] The UK government's January 2026 also suggests that workers in highly exposed occupations differ materially depending on whether their roles entail higher or lower complementarity with AI.[21] OECD reports emphasise that AI human expertise remains an important determinant of value creation, even under broad AI diffusion, with training and worker consultation being associated with better outcomes.[22]

The empirical picture reflects the complexity involved in predicting the impact of AI on jobs. A study of 5,172 customer support agents found a 15% average productivity increase following AI adoption, with larger gains among less experienced workers, suggesting that AI can raise output without reducing headcount and may disproportionately benefit those earlier in their careers.[23] A 2026 cross-country executive survey reported widespread firm-level AI adoption, while also finding that most surveyed firms reported stable employment and productivity over the prior three years.[24] Aggregate data show unemployment among highly exposed workers has remained stable.[25] Adoption is spreading quickly while broad labour market displacement remains difficult to detect until data are disaggregated by age and occupational exposure.[26]

One area where impacts are more clearly observed is hiring of younger workers. A recent US study, for instance, shows that workers aged 22 to 25 in the most AI-exposed occupations experienced a 16% relative decline in employment from late 2022 to September 2025, against growth of 6% to 9% among older workers in the same occupations.[27] A 2026 US Federal Reserve note finds job postings in industries and firms with higher AI adoption have held steady, while pointing to evidence of entry-level employment declines in occupations where AI mainly automates tasks.[28] This formative work problem deserves particular attention because its human capital consequences accumulate slowly and are difficult to reverse once in place. Early career work has historically produced experienced professionals who would over time move into leadership and expert roles. Yet, the work through which junior professionals develop technical skills and professional networks is precisely the work that AI systems are increasingly capable of performing well.[29] Where firms reduce early-career work and leave developmental pathways unchanged as they deploy AI internally, they may gain near-term efficiency while weakening the developmental pipeline of experienced practitioners. Moreover, the capacity of organisations to exercise genuine oversight of AI systems depends on a supply of experienced judgement that firms may inadvertently be eroding. It remains to be seen whether the effect on early-career work is a temporary trend or whether it translates into a longer-term risk for investors. **However, investors should ask companies to evidence that AI adoption is accompanied by robust workforce planning and organizational and cultural change management.**

20. International Labour Organization. 2025. Generative AI and Jobs: A Refined Global Index of Occupational Exposure. Available at: https://www.ilo.org/publications/generative-ai-and-jobs-refined-global-index-occupational-exposure
21. UK Government. 2026. Assessment of AI Capabilities and the Impact on the UK Labour Market. Published 28 January 2026. Available at: https://www.gov.uk/government/publications/assessment-of-ai-capabilities-and-the-impact-on-the-uk-labour-market/assessment-of-ai-capabilities-and-the-impact-on-the-uk-labour-market
22. OECD. 2025. OECD Employment Outlook 2025. Sections on AI, work, job quality, training and worker consultation. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/07/oecd-employment-outlook-2025_5345f034/194a947b-en.pdf
23. Brynjolfsson, E., Li, D., & Raymond, L. 2025. Generative AI at work. The Quarterly Journal of Economics, 140(2), 889-942.
24. Yotzov et al., 2026. Firm Data on AI. NBER Working Paper No. 34836 / Atlanta Fed Working Paper, 24 March 2026. Available at: https://www.nber.org/papers/w34836
25. The Budget Lab at Yale. 2026. "Tracking the Impact of AI on the Labor Market." Updated April 2026. https://budgetlab.yale.edu/research/tracking-impact-ai-labor-market.
26. Liu, J & Webber, D. 2026. "AI Adoption and Firms' Job-Posting Behavior." FEDS Notes. Board of Governors of the Federal Reserve System, 27 March 2026. DOI: 10.17016/2380-7172.4026. The Budget Lab at Yale. 2026, op cit.
27. Brynjolfsson, E., Chandar, B., & Chen, R. (2025). Canaries in the Coal Mine?': Six Facts about the Recent Employment Effects of Artificial Intelligence.
28. Liu, J & Webber, D. 2026. "AI Adoption and Firms' Job-Posting Behavior." FEDS Notes. Board of Governors of the Federal Reserve System, 27 March 2026. DOI: 10.17016/2380-7172.4026.
29. Brynjolfsson, E., Li, D., & Raymond, L. 2025. "Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence." Working paper. Stanford Digital Economy Lab, November 2025. Available at https://digitaleconomy.stanford.edu/publications/canaries-in-the-coal-mine/.

Governance of AI in the workplace

An ongoing social issue in AI diffusion is how (and how much) employees will interact with new AI tools and systems, as well as the extent to which these tools will be used to manage labour. Many companies are deploying AI faster than they are redesigning work around it, and that **governance lag often becomes visible before firms have fully settled how AI should be used, making it an important area for investor engagement.**[30]

Two concepts are useful to consider here: workforce capability concerns whether workers can supervise and use AI well, whereas worker voice is related to whether employees have a meaningful say in deployment. The latter is an important AI governance quality indicator, as employees are likely to be the first to notice where AI systems do not function as intended.[31] Where employees turn to external AI tools because internal systems are slow, poorly aligned with actual work, or inadequately trusted to support it, a phenomenon known as "shadow AI", formal governance may diverge from how AI is being used in practice. MIT's Project NANDA reports that workers at over 90% of surveyed companies use personal AI tools for work, while fewer than 40% of those companies have purchased official subscriptions.[32] A company can therefore appear robust in policy terms while actual use is already becoming harder to monitor and govern. Firms that provide credible avenues for worker feedback on AI deployment are better placed to catch governance failures before they become embedded in organisational processes.[33]

Governance shortcomings and poor management of organizational change can also lead to process redesign whereby employees will face an extensive burden of reviewing AI outputs, which in itself can be unproductive. A 2025 study on generative AI and critical thinking among knowledge workers finds that higher confidence in generative AI is associated with lower reported critical thinking effort.[34] Human involvement in AI-assisted processes therefore gives only a partial picture of oversight quality. In stronger settings, AI supports judgement by removing repetitive work and giving workers more room to focus on what requires interpretation and other value-added tasks. In settings with weaker governance, speed of AI adoption rises while scrutiny weakens, and workers retain responsibility for outcomes with less room to manage the risk well.

The same issue also appears in how productivity gains are leveraged. If this is not properly considered, workers may have carry accountability for outputs while the firm keeps the gain, which can result in disengagement. EY's 2025 US AI Pulse Survey suggests a more constructive possibility, with many firms reporting the channeling of AI-related gains into further capability development, retraining and related investment rather than immediate workforce reduction.[35] **This points to the need for investors to assess what happens within that broad category of reinvestment.**

The amount of work involved in leveraging and reviewing AI output may evolve through job redesign as firms adapt and models improve, leading to more reliable outputs. Developers can influence these outcomes through the default workflow assumptions they make and the implementation guidance they provide to enterprise customers. **Investors may therefore ask developers about their own workforce impacts and for clearer public disclosure on intended enterprise use cases, where meaningful human oversight is expected, and what guidance is given to customers on safe deployment in labour-sensitive settings.**

30. Yotzov, I. et al. 2026. Firm Data on AI. NBER Working Paper No. 34836 / Atlanta Fed Working Paper, 24 March 2026. Available at: https://www.nber.org/papers/w34836
31. Trades Union Congress. 2025. General Council Report 2025, section on AI at work and algorithmic management. Available at: https://www.tuc.org.uk/research-analysis/reports/general-council-report-2025?page=3
32. MIT Project NANDA. 2025. The GenAI Divide: State of AI in Business 2025. Available at: https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf
33. See OECD. 2025 and Trades Union Congress. 2025, op cit.
34. Lee, H.-P. et al. 2025. "The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects from a Survey of Knowledge Workers." Microsoft Research / CHI 2025. Available at: https://dl.acm.org/doi/10.1145/3706598.3713778
35. EY. 2025. "AI-driven productivity is fueling reinvestment over workforce reductions." Fourth EY US AI Pulse Survey, 9 December 2025. Available at: https://www.ey.com/en_us/newsroom/2025/12/ai-driven-productivity-is-fueling-reinvestment-over-workforce-reductions

Algorithmic management and AI-driven transformation of work

Algorithmic management encompasses a range of digitally enabled managerial practices, including automated assignment of tasks and shifts, performance monitoring, evaluation of productivity against benchmarks, and automated or semi-automated decisions about pay, promotion and termination. A 2025 European Parliament study estimates that 42.3% of EU workers are already exposed to algorithmic management, with exposure potentially rising to 55.5% over the medium term.[36] Human Rights Watch's May 2025 reportdocumented algorithmic wage management and labour exploitation in platform work across the United States, finding that platform companies use algorithmic systems to manage worker output and compensation while classifying workers as independent contractors to avoid the obligations of employment.[37]

The spread of algorithmic management into traditional employment relationships materially changes the character of work for millions of employees. The governance problem of algorithmic management lies in the absence of meaningful limits around these systems. As AI enters performance and management processes, workers become more legible to automated decision-making while their ability to question outcomes diminishes. Where decisions affecting pay or career progression are shaped by algorithmic systems without adequate explanation or appeal, workers lose practical access to the rights that employment law nominally provides. This makes algorithmic management another area of increased regulatory scrutiny, with obligations that global deployers are still often ill-equipped to navigate consistently.[38]

3. Downstream issues

Consumer harm and access to services

AI systems are now widely embedded in decisions that determine access to vital services for significant populations of customers, including healthcare and finance. This application of AI often raises concerns about discrimination and has attracted regulatory scrutiny. The EU AI Act classifies many of these applications as high risk, creating documentation and human oversight obligations.[39] In the US, existing anti-discrimination law has been confirmed to apply to automated decision systems regardless of whether discrimination is intentional.[40]

Proxy discrimination in AI-assisted decisions is difficult to detect through standard compliance review. A simple variable such as postcode can appear neutral while functioning as a close proxy for race or income in a given decision context, producing discriminatory outcomes at scale across decision populations that comply with the letter of the law at the individual level. Many firms assess individual decisions more readily than aggregate patterns across decision populations, increasing risks of systematic bias seeping into otherwise compliant processes.[41]

The consequences of governance shortcomings in these settings extend well beyond individual harm. Public failures of automated and algorithmic decision systems, including the Robodebt scheme in Australia and the Dutch childcare benefits scandal, illustrate how high-volume systems can produce population-scale harm before they are identified and corrected.[42] Robodebt wrongly issued hundreds of thousands of debt demands and was subsequently the subject of a royal commission. The Dutch childcare benefits

36. European Parliamentary Research Service. 2025. Digitalisation, Artificial Intelligence and Algorithmic Management in the Workplace: Shaping the Future of Work. EPRS_STU(2025)774670, 24 October 2025. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774670/EPRS_STU(2025)774670_EN.pdf
37. Human Rights Watch. 2025. The Gig Trap: Algorithmic, Wage and Labor Exploitation in Platform Work in the US. Available at: https://www.hrw.org/report/2025/05/12/the-gig-trap/algorithmic-wage-and-labor-exploitation-in-platform-work-in-the-us
38. European Union. 2024. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence, Artificial Intelligence Act. Colorado General Assembly. 2024. Colorado Artificial Intelligence Act, SB24-205. New York City Department of Consumer and Worker Protection. 2023. Local Law 144 of 2021: Automated Employment Decision Tools.
39. European Union. 2024. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence, Artificial Intelligence Act.
40. Federal enforcement capacity has since contracted substantially, and state-level action, particularly in California and New York, has become the primary enforcement mechanism.
41. Consumer Financial Protection Bureau, Equal Employment Opportunity Commission, Federal Trade Commission and Department of Justice. 2023. Joint Statement on Enforcement Efforts Against Discrimination and Bias in Automated Systems. Available at: https://www.ftc.gov/system/files/ftc_gov/pdf/EEOC-CRT-FTC-CFPB-AI-Joint-Statement%28final%29.pdf
42. Royal Commission into the Robodebt Scheme. 2023. Report of the Royal Commission into the Robodebt Scheme. Available at: https://robodebt.royalcommission.gov.au/publications/report

scandal involved automated fraud detection that disproportionately flagged families of dual nationality and resulted in state compensation running to billions of euros.[43]

Where AI systems systematically underserve or misprice risk for particular demographic groups, this can deepen of existing disparities in access to financial and healthcare services, which in turn can create negative systemic impacts for the economy. Investors should assess whether companies monitor AI-assisted decisions at the population level as well as the individual level, and whether affected consumers retain practical routes to explanation and remediation when outcomes are adverse.

Child rights

Child safety has become one of the most rapidly developing areas of AI-related legal and regulatory risk, driven by a convergence of social media litigation, AI companion harms and a broadening legislative response that is now moving faster than many companies anticipated.

One of the most significant legal developments of 2026 is the verdict in K.G.M. v. Meta and YouTube, handed down by a California jury in March 2026. The jury found Meta and Google liable in a social media harm case involving Instagram and YouTube, ordering the companies to pay a combined US$6 million in damages. A New Mexico jury separately ordered Meta to pay US$375 million after finding the company misled users about platform safety and enabled child exploitation on its platforms. Related personal injury cases and school district claims are pending nationwide, consolidated in federal and state proceedings.[44] The cases focus primarily on platform design and algorithmic recommendation systems; AI-driven features are directly implicated in the alleged harms, and regulators are increasingly treating the two as inseparable. US lawmakers introduced large numbers of state children's online safety bills in 2026, targeting how companies collect data from minors and expose them to harmful content, predatory behaviour and addictive features.

The newest regulatory frontier is AI companion chatbots. A new law in California, effective January 2026, was the first in the United States to regulate companion chatbot operators, requiring safety protocols around interactions with minors, including regular reminders that the user is interacting with an AI system,[45] with several other US states having considered or introduced comparable measures.[46] The US context is instructive as it is home to many of the major developers and other large technology companies for whom these issues are highly material. However, this regulatory trend is global and accelerating. China's Cyberspace Administration draft 2026 rules require virtual human content to be clearly labelled and prohibit services that create addictive or emotionally manipulative interactions with underage users.[47] In the United Kingdom, the government recently announced that it would move to bring AI chatbot providers within the illegal content duties of the Online Safety Act.[48] Indonesia has also moved to restrict access by childreno high-risk digital platforms.[49]

AI-generated child sexual abuse material (CSAM) represents a related and more severe escalation of these risks. National Center for Missing & Exploited Children reported a 1,325% increase in online exploitation reports involving generative AI, from 4,700 in 2023 to 67,000 in 2024. The Internet Watch Foundation separately recorded a staggering increase in actionable reports containing AI-generated child sexual abuse imagery in 2024, up 380% since 2023.[50] Synthetic media harms more broadly have grown as a share of recorded AI incidents, rising 2.5 times between 2022 and 2025.[51]

Child safety increasingly draws broad public and policymaker support, with clearer regulatory direction of regulatory. The compliance burden for companies will likely continue to widen as legal risks increase, and **investors should assess whether companies are building governance at the pace that this trajectory demands.**

43. Berends, S. 2021. "Dutch child benefit scandal: origin and latest developments." ESPN Flash Report 2021/51. European Social Policy Network / European Commission, July 2021. Available at https://ec.europa.eu/social/BlobServlet?docId=24723&langId=en.
44. For a summary, see: https://www.reuters.com/legal/litigation/jury-reaches-verdict-meta-google-trial-social-media-addiction-2026-03-25/
45. California Legislature. 2025. SB 243, Companion Chatbots.
46. National Conference of State Legislatures. AI legislation tracker. Available at: https://www.ncsl.org/financial-services/artificial-intelligence-legislation-database
47. Reuters. 2026. "China moves to regulate digital humans, bans addictive services for children." 3 April 2026. Available at: https://www.reuters.com/world/china/china-moves-regulate-digital-humans-bans-addictive-services-children-2026-04-03/
48. UK Government. 2026. "PM: No platform gets a free pass: Government takes action to keep children safe online." 15 February 2026.Available at: https://www.gov.uk/government/news/pm-no-platform-gets-a-free-pass-government-takes-action-to-keep-children-safe-online
49. Associated Press. 2026. Indonesia urges social media platforms to disclose the number of accounts closed for users under 16. 29 April 2026. Available at: https://apnews.com/article/indonesia-social-media-children-under-16-39630c776f947652cde619ad4ae56627
50. National Center for Missing & Exploited Children. 2025. CyberTipline Data. Available at https://missingkids.org/gethelpnow/cybertipline/cybertiplinedata. Internet Watch Foundation. 2024. Annual Data & Insights Report 2024: AI-generated Child Sexual Abuse. Available at: https://www.iwf.org.uk/annual-data-insights-report-2024/data-and-insights/ai-generated-child-sexual-abuse/

Defence and dual use

AI-enabled targeting, autonomous drone functions and counter-drone systems have now moved into active military use, raising fundamental questions about how the international humanitarian law principles apply when targeting decisions are made or supported by automated systems. Where targeting decisions produce lethal harm and accountability cannot be assigned to an identifiable party, the result is a responsibility gap that translates into legal exposure and procurement restrictions for companies involved in developing or integrating such systems.[52]

An even more immediate investor concern is the dual-use technology problem. Computer vision systems developed for autonomous vehicles can be repurposed for drone targeting, whilst natural language processing developed for enterprise productivity can be integrated into intelligence and surveillance systems. Investors in civilian AI companies may therefore carry indirect exposure to military AI applications without being captured by the governance frameworks typically applied to defense investments. Although this issue spans the value chain in ways that extend beyond this paper's primary focus on corporate-level social governance, it is included here because the investor governance question of whether civilian AI companies have responsible use policies governing military and law enforcement applications and enforce them through customer screening and senior-level accountability is one that deployers as well as developers need to answer.[53]

Dual-use decisions can generate reputational and operational risk where governance is unclear, as illustrated by employee whistleblowing and organising in firms involved in military and surveillance contracts. The harm pathways extend beyond the use of weapons systems alone. UNICEF's 2025 guidance explicitly addresses AI use in armed conflict and its implications for children, noting that AI-enhanced targeting and amplification tools may enable recruitment, grooming or manipulation of children in conflict settings, and that autonomous or AI-enabled systems operating in civilian-populated areas increase harm to children.[54] Investors may therefore start to consider a child rights dimension to dual-use governance that existing responsible AI frameworks rarely address.

Community rights and data centre infrastructure

The physical buildout of AI has become one of the largest capital allocation stories in the global economy. Data centre investment reached US$770 billion in 2025, exceeding upstream oil and gas in the same year and representing a more than sixfold increase from US$112 billion in 2020.[55] The scale is concentrated among a remarkably small number of companies, and the infrastructure they are building lands on local grids, land and water systems. This has led to growing community opposition which is emerging as a material constraint on the AI infrastructure buildout. In Q2 2025 alone, nearly US$98 billion in proposed investment was associated with projects blocked or delayed amid local opposition, more than the total disruption it had tracked in all earlier quarters since 2023.[56] Sightline Climate's 2026 outlook similarly argues that attrition in the pipeline is becoming material: it estimates that 30 to 50% of large data centre capacity slated for 2026 may be delayed, and notes that only about 5 GW of the 16 GW planned for 2026 is already under construction.[57]

Opposition is bipartisan and increasingly coordinated. In 2025, US officials paused data centre approvals in Maryland, and in Georgia eight counties and cities adopted temporary moratoria during 2025 while drafting data-centre-specific rules. In Maine, legislators considered establishing a data centre coordination council to study impacts before projects proceed. These examples suggest that communities that once competed for hyperscale investment are increasingly seeking time, leverage and procedural control before projects proceed.[58]

51. OECD. 2026. Trends in AI Incidents and Hazards Reported by the Media. Available at: https://www.oecd.org/en/publications/trends-in-ai-incidents-and-hazards-reported-by-the-media_4f5ff43c-en.html
52. International Committee of the Red Cross. 2024. Autonomous Weapons. ICRC Law and Policy. Available at https://www.icrc.org/en/law-and-policy/autonomous-weapons. United Nations Office for Disarmament Affairs. Lethal Autonomous Weapon Systems. Available at https://disarmament.unoda.org/en/our-work/emerging-challenges/lethal-autonomous-weapon-systems.
53. See also International AI Safety Report. 2026. Op cit.
54. Equidem. 2025. Scroll. Click. Suffer: The Hidden Human Cost of Content Moderation and Data Labelling. 28 May 2025. Available at: https://equidem.org/reports/scroll-click-suffer-the-hidden-human-cost-of-content-moderation-and-data-labelling/
55. Rystad Energy. 2026. "Putting Things in Perspective: Data Center Investments Now on Par with Renewables." March 2026. Available at: https://www.rystadenergy.com/insights/putting-things-in-perspective-data-center-investments-now-on-par-with-renewables
56. Data Center Watch. 2025. Q2 2025 Data Center Opposition Update. Available at: https://datacenterwatch.org.
57. Sightline Climate. 2026. Data Centre Market Outlook 2026. Available at: https://sightlineclimate.com.
58. See https://www.nbcnews.com/politics/politics-news/reining-data-centers-sparks-rare-bipartisanship-statehouses-rcna262990 for a summary.

The resource demands on host communities explain the depth of opposition. Data centres draw heavily on local power grids and, in several grid regions, contribute to electricity cost increases that fall hardest on lower-income households. Carnegie Mellon estimates that average US electricity generation costs could rise by 8% by 2030 due to data centre and cryptocurrency demand, with grid upgrade costs in some regions socialised across wider utility customer bases.[59] Water demand for cooling creates direct competition with agricultural and residential users. In water-stressed regions, these demands intersect with existing social and economic vulnerability. The siting of hyperscaler facilities can also generate persistent noise, alter land use and change the character of communities that did not anticipate becoming data centre hubs. In some jurisdictions, fiscal arrangements leave host communities bearing infrastructure burdens without receiving the level of public revenue that headline investment figures imply.

A significant driver of opposition is limited transparency on how many data centre projects are planned and disclosed. Sightline notes that roughly a quarter of the 140 projects expected to come online by 2030 have not disclosed how they plan to secure power, despite the centrality of power sourcing to both feasibility and community impact.[60] Limited disclosure to local stakeholders on project ownership and plans on powering data centres is increasingly leading to project delays and attrition. In the Netherlands, Microsoft and Google were reported in 2026 to be withholding data centre energy use data despite EU transparency rules. In Spain's Aragón region, water use requests associated with hyperscale expansion have heightened local concern about resource burden allocation. Where siting decisions proceed without adequate community consultation, the result is a procedural legitimacy problem that generates opposition and permitting delays.[61]

The communities bearing the primary impacts are also frequently not the communities that benefit most from AI services or from the wider economic gains of the AI industry. Electricity bill increases, for instance, fall proportionally harder on lower-income households that already spend a higher share of income on energy. In December 2025, more than 230 environmental organisations called for a halt to new US data centre construction, citing threats to energy affordability, water resources and wider environmental security.[62]

Overall, investors should seek corporate disclosure that goes beyond aggregate resource figures towards site-level data that allows meaningful assessment of where costs fall, whether local concerns have influenced project design, as well as whether community engagement is built into project development from the outset rather than appended after decisions have been made.

III. AI governance in practice: evaluations, audits and risk mitigation

The social risks discussed require companies to have the appropriate structures to identify and manage those risks before they translate into harm. These AI governance processes take on various forms. Safety evaluations test how a model or system behaves in practice: whether it produces harmful outputs, performs unevenly across groups, or fails under realistic conditions. Audits examine whether the governance process around a system is working as claimed. The two are related but not the same. A company can, for instance, run credible technical evaluations while maintaining weak organisational oversight, but it can also commission an audit scoped so narrowly as to provide investors with limited assurance. When companies say they have reviewed their AI, the productive questions are which process was used, against what criteria, by whom and with what findings.

59. See https://www.cmu.edu/work-that-matters/energy-innovation/data-center-growth-could-increase-electricity-bills
60. Sightline Climate. 2026. Data Centre Market Outlook 2026. Op cit.
61. Environmental Defense Fund and Nuveen. 2026. Decoding Data Centers: Sustainability Due Diligence Across the Value Chain. Available at: https://business.edf.org/insights/decoding-data-centers-sustainability-due-diligence-across-the-value-chain/
Data Center Watch. 2025. Q2 2025 update on data centre development risk and local opposition, available at: https://www.datacenterwatch.org; see also Wired, "The Data Center Resistance Has Arrived," 14 November 2025. Available at: https://www.wired.com/story/the-data-center-resistance-has-arrived/. Axios, "Local opposition creates roadblocks for AI boom," 24 February 2026. Available at: https://www.axios.com/2026/02/24/ai-data-centers-energy-bills
62. The Guardian. 2025. More than 200 environmental groups demand halt to new US datacenters. 8 December 2025. Available at: https://www.theguardian.com/us-news/2025/dec/08/us-data-centers

Delivering on these processes also requires solid internal ownership. A robust governance function should have the authority to delay high-risk deployment or trigger review when risks change or controls prove inadequate. Advisory functions with no authority to delay or restrict deployment are a weak governance indicator.

Table 1 sets out the main AI governance processes and what each tells investors.[63]

The evidence investors should expect differs substantially between developers and deployers. The relevant questions shift from model behaviour in the abstract for developers to what happens when AI is inserted into a specific workflow and user population for deployers. Figure 1 sets out what strong and weak evidence looks like for each, and between the two.

Table 1. The AI governance architecture: core processes and evidence

PROCESS	CORE QUESTION	WHAT IT TELLS INVESTORS
AI safety evaluation	Does this system behave acceptably in this context?	Whether the company assesses system-level AI risk beyond performance metrics, covering bias, privacy, misuse and realistic failure conditions.
AI audit / assurance review	Is governance around this system credible and functioning?	AI governance quality and traceability: whether process claims are backed by evidence.
Human rights impact assessment	Whose rights could be affected and what is the company doing about it?	Whether AI impacts on workers, users and communities are mapped, affected groups consulted, and remedy pathways in place.
Risk-tiering / algorithmic impact assessment	How risky is this use case and what controls should apply?	Whether the company distinguishes between higher- and lower-risk AI applications and applies proportionate governance to each.
Enterprise AI risk management	How does the company connect governance across the deployment lifecycle?	Whether AI intake, evaluation, approval, monitoring and incident management operate as a connected process or as isolated exercises.

Key sources: NIST AI Risk Management Framework; NIST Generative AI Profile; OECD Due Diligence Guidance for Responsible AI; UNESCO Ethical Impact Assessment; Singapore Model AI Governance Framework for Agentic AI; ISO/IEC 42001.

63. The MIT AI Risk Initiative has compiled a publicly accessible taxonomy of 831 AI risk mitigations across four categories, which provides a useful map for investors tracing specific mitigation types to governance expectations. Saeri, Alexander K., et al. 2025. Mapping AI Risk Mitigations: Evidence Scan and Preliminary AI Risk Mitigation Taxonomy. MIT AI Risk Initiative / arXiv.

Figure 3. What investors should look for in AI governance disclosures

STRONG AND WEAK EVIDENCE ACROSS THE DEVELOPER-DEPLOYER CHAIN

AI DEVELOPERS | **AI DEPLOYERS**

Feedback ⟷ Documentation ☑

AREA	✕ STRONG EVIDENCE	✓ WEAK EVIDENCE
Model and System Documentation	Detailed documentation with clearly stated intended and prohibited uses; limitations described concretely; evaluation categories named; updated when sxhange materially.	Generic capability summaries; benchmark results without operational caveats; no indication of what was tested, found or restricted.
Pre-release safety and misuse evaluation	Testing categories described, including harmful outputs, bias, privacy and robustness; testing for how the system can be manipulated or misused is referenced; findings and mitigations disclosed.	Evaluation claimed without scope, methodology or findings; safety described in aspirational terms only.
Data governance and provenance	Training data sourcing disclosed; known quality issues or dataset limitations documented; prohibited sources identified	No disclosure of data sourcing practices, quality controls or known limitations.
Limitation disclosure and use restrictions	Failure modes stated; performance limits in specific contexts described; restricted or out-of-scope uses communicated to downstream integrators.	Performance described without limitations; no guidance on inappropriate or high-risk deployment contexts.
Support for downstream users	Documentation provided to integrators covering safe use, known risks and appropriate governance; implementation guidance available.	Models available without documentation on limitations, safeguards or downstream governance expectations.

AREA	✕ STRONG EVIDENCE	✓ WEAK EVIDENCE
Use case scoping and inventory	Material AI use cases disclosed; distinction made between human-assisted and more autonomous applications; prohibited or restricted uses stated.	AI presence acknowledged without disclosure of specific uses, risk classification or restrictions.
Testing in deployment context	Testing conducted in the actual deployment context and user population; bias, privacy and safety evaluated in the specific workflow rather than only at model level.	Reliance on developer testing without independent assessment of the specific workflow and use case.
Criteria for human review	Criteria for when human review is required are documented; escalation routes and override rights exist; staff equipped to challenge outputs in practice.	Human involvement asserted without criteria, escalation routes or evidence that meaningful review is possible under operational conditions.
Vendor due diligence	Third-party AI systems assessed against company's own governance standards; ongoing monitoring of vendor-supplied components; procurement used as governance lever.	Vendor reputation treated as a substitute for in-context evaluation; no disclosure of how third-party systems are assessed or monitored.
Incident monitoring	Incident capture, complaint handling and escalation processes described; changes in model behaviour tracked; process for restricting/withdrawing systems documented.	Deployment treated as a final governance event; no disclosed process for monitoring, complaints or remediation after deployment.
Routes to remedy and challenge	Employees and customers have accessible channels for raising concerns; AI-assisted decisions can be reconstructed and reviewed; escalation to human decision-makers is available in practice	Complaints channel exists formally but AI-assisted decisions cannot be challenged in practice.
Whistleblower reporting and protection	Named reporting channels through which employees can raise AI-related concerns confidentially; non-retaliation commitments covering AI safety disclosures; governance documentation confirming that concerns can be escalated independently of commercial teams	General ethics hotlines with no indication that AI-specific concerns are in scope, or that reporters are protected from retaliation.

The processes described above address AI risk primarily through a technical and operational lens and should be complemented by human rights due diligence and human rights impact assessments. These processes are designed to address questions that sit outside the scope of system-level evaluations, such as whether those impacts are acceptable, whether affected groups have been meaningfully consulted and whether remedy is accessible in practice. The question of access to remedy is particularly challenging to address: in many consequential uses, deployers receive complaints first, yet the relevant safeguards and design choices may sit upstream with the developer or vendor. For investors, an important question is whether responsibility for incident response and remediation is workable across the developer-deployer chain.[64]

IV. AI governance frameworks: convergence and gaps

Assessing whether companies are managing social risks of AI adequately requires a clear view of what good governance looks like. This section identifies where convergence across regulatory instruments and voluntary standards gives investors a useful reference point for answering that question, and where gaps in operational metrics leave the assessment question genuinely open.

Our review of cross-sector frameworks spanning regulatory instruments, voluntary standards and sector guidance across multiple jurisdictions finds broad alignment on a common set of operational requirements.[66] The developer and deployer distinction is increasingly explicit, with regulatory and normative guidance generally distinguishing upstream model governance from the operational accountability that sits with deploying companies. Where existing frameworks remain weakest is on the questions most central to the risks identified in the preceding sections: whether human oversight of AI is genuinely meaningful under operational conditions, and whether affected workers, users and communities have workable routes to challenge outcomes that affect them and access remedy. Frameworks provide less guidance on how to assess whether grievance and remedy processes for AI specifically are functioning in practice. The absence of standardised operational metrics in these areas means investors need to engage with companies proactively to distinguish between formal compliance and effective control. The weight to place on specific governance expectations will depend in part on the pace at which AI capability and adoption develop, a question the following section examines through scenario analysis.

64. See BSR: https://www.bsr.org/files/BSR-Remedy-for-Generative-AI-Related-Harms.pdf

65. François, C.et al. 2026. "Why AI Model Cards Are an Urgent Necessity for Child Safety." Tech Policy Press, 2 April 2026. Available at https://techpolicy.press/why-ai-model-cards-are-an-urgent-necessity-for-child-safety.

66. Representative frameworks reviewed: Core AI governance and risk management: NIST AI Risk Management Framework; NIST Generative AI Profile; OECD AI Principles; OECD Due Diligence Guidance for Responsible AI; ISO/IEC 42001; ISO/IEC 23894; UNESCO Recommendation on the Ethics of AI; UNESCO Ethical Impact Assessment; UNESCO Readiness Assessment Methodology. Regulatory and policy instruments: EU AI Act; EU General-Purpose AI Code of Practice; Council of Europe Framework Convention on AI; UK principles-based AI regulatory framework; Singapore Model AI Governance Framework, AI Verify and Model AI Governance Framework for Agentic AI; Japan AI Guidelines for Business; Canada Voluntary Code of Conduct for Advanced Generative AI Systems; Australia voluntary AI safety and assurance materials; Colorado AI Act; New York City Local Law 144. Frontier model and developer sources: G7 Hiroshima AI Process International Code of Conduct; Seoul Frontier AI Safety Commitments; Frontier Model Forum; International AI Safety Report 2026; METR; SaferAI Risk Management Ratings; Future of Life Institute AI Safety Index; Foundation Model Transparency Index. Rights, workforce and vulnerable-user sources: UN Guiding Principles on Business and Human Rights; OHCHR B-Tech Project; UNICEF Guidance on AI and Children; ILO sources on AI, work and employment; WHO ethics and governance guidance for AI in health; Ranking Digital Rights; World Benchmarking Alliance AI accountability work. Assurance, audit and transparency sources: AI Verify and Global AI Assurance Sandbox; MLCommons AILuminate; ISACA AI audit materials; ForHumanity audit and certification criteria; Ada Lovelace Institute work on AI assurance; Partnership on AI responsible deployment materials; C2PA Content Credentials. Monitoring and indicator sources: OECD AI Incidents Monitor; Stanford HAI AI Index; IEA Energy and AI; MIT AI Risk Initiative; Yale Budget Lab labour-market monitoring; NCSL AI legislation tracker; IAPP AI governance tracker; Synergy Research Group cloud market data.

Table 2. Areas of convergence across AI governance frameworks

AREA OF CONVERGENCE ACROSS FRAMEWORKS AND REGULATIONS	POINTS OF AGREEMENT	WEAKER OR LESS DEVELOPED AREA
Governance model	AI governance is increasingly framed as an operating model: inventory, risk classification, evaluation, approval, documentation, monitoring and review.	Outcome evidence is less developed than process design.
Risk and context	Expectations are risk based and use-case specific, with stronger controls for consequential uses.	Firms still describe use cases unevenly, which limits comparability.
Oversight and control	Frameworks expect named responsibility, human oversight, escalation routes and the ability to restrict or withdraw systems.	Oversight is often clearer in principle than in day-to-day operating conditions.
Testing and documentation	Good practice now includes testing beyond accuracy, plus documentation on uses, limits, approvals and traceability.	Public disclosure is still often too general to show how strong control really is.
Monitoring and remedy	Governance is expected to continue after deployment through incident tracking, complaints, review and control updates.	Remedy and contestability remain less developed than the rest of the governance architecture.
Roles and assurance	There is clearer separation between developer and deployer responsibilities, and growing emphasis on audit and assurance.	External assurance remains uneven, especially for ordinary deployers.

Agentic AI as an emerging governance frontier

Commercial deployment of agentic AI is expanding rapidly across many sectors, and the governance question it raises extends beyond whether a model produces acceptable outputs to whether a system of interacting components remains controllable, attributable and correctable when errors can propagate through automated workflows before any human reviews them. AI agents and automated workflows already outnumber human identities in enterprise environments by ratios exceeding 80 to 1, yet integrated governance frameworks for these systems remain largely absent[67].

Agentic AI shifts questions for deployers from reviewing outputs to defining, before deployment begins, what tasks an agent may perform and what limits apply, because in automated multi-step workflows the point at which a consequential decision is made may not be visible to a human reviewer at all. Agent actions, including sending communications, modifying records and triggering transactions, can propagate before review is possible and may be difficult to reverse; where multiple agents work in sequence, problems can compound across the chain in ways that make it difficult to identify what went wrong and who bears responsibility. Investors should ask whether companies have defined task boundaries and approval gates before agentic deployment, and whether their governance frameworks address the accountability gaps that multi-agent architectures create.

67. Full reference for footnote: Kurtz, A., & Krawiecka, K. 2026. Who Governs the Machine? A Machine Identity Governance Taxonomy (MIGT) for AI Systems Operating Across Enterprise and Geopolitical Boundaries. arXiv preprint arXiv:2604.06148.

V. Scenarios and adaptive stewardship priorities

AI continues to evolve at a rapid pace, and investors cannot assume a single stable diffusion trajectory against which to judge whether companies are managing risks well. The pace of capability improvement will shape how quickly AI moves into consequential uses: thus, the pattern of enterprise adoption will likely determine whether risks remain concentrated in specialist settings or spread through ordinary workflows, whereas market concentration will influence how much leverage deployers retain over the systems on which they depend.

Scenario analysis is therefore necessary complement to stewardship on responsible AI. Narayanan and Kapoor's (2025) treatment of AI as a normal technology is a useful starting point: even transformative technologies produce their effects through applications, adoption and diffusion, and diffusion is shaped by the speed at which organisations and institutions adapt.[68] That framing makes deployment quality the central stewardship variable regardless of which capability trajectory materialises, and it cautions against treating any single scenario as the fixed baseline.

The OECD's 2026 analysis of possible AI trajectories through 2030 provides one of the most comprehensive scenario frameworks available at the time of writing. It identifies four plausible AI diffusion scenarios: progress stalls at roughly current capability levels; progress slows as gains continue but at a diminished rate; progress continues at roughly the current rapid pace; or progress accelerates beyond current rates. The analysis concludes that available evidence is insufficient to discount any of these scenarios, and that experts consulted in its preparation expressed high uncertainty and low confidence in their ability to predict the rate of AI progress through 2030.[69]

So far, evidence suggests that AI diffusion advances rapidly in some domains and more gradually in others. Stanford HAI's 2026 AI Index found that generative AI reached 53% population adoption within three years, and organisational adoption rose to 88%, yet deployment of AI agents remains in single digits across almost all business functions. Capability progress is similarly uneven: AI systems can now achieve gold medal performance in competitive mathematics, yet a leading model reads analogue clocks correctly only 50.1% of the time. Agent performance on ordinary computer tasks has improved from roughly 12% to approximately 66% task completion, while still failing in roughly one in three attempts.

The governance picture is correspondingly complex. Documented AI incidents rose from 233 in 2024 to 362 in 2025, responsible AI benchmark reporting remains patchy, and 80 of 95 notable models released in 2025 were released with no training code disclosed. Public confidence lags expert optimism markedly: in 2026, 73% of experts expected AI to improve how people do their jobs, against 23% of the public, and only 31% in the United States said they trusted government to regulate AI effectively.[70]

Scenarios can help investors distinguish between risks that require engagement now and risks that become more material if deployment accelerates, concentrates or moves into higher-consequence settings. The fundamental stewardship question is consistent across all paths: can the company evidence a governed AI lifecycle from inventory and approval through monitoring and correction? The weight of that question, and the consequences of an inadequate answer, increase as the capability trajectory moves towards the upper end of the plausible range. Some stewardship priorities are stable across all scenarios, including deployment discipline, vendor governance, worker voice, and credible routes to remedy, whereas others are path-dependent. The figure below maps the principal social ESG risk channel, the shift in stewardship emphasis, and the observable signals that would indicate a given trajectory is materialising under each of the four OECD scenario classes.

68. Narayanan, A. & Kapoor, S. 2025. "AI as Normal Technology: An alternative to the vision of AI as a potential superintelligence". Knight First Amendment Institute. Available at: https://kfai-documents.s3.amazonaws.com/documents/0ee1da899a/AI-as-Normal-Technology---Narayanan---Kapoor-Final.pdf
69. OECD. 2026. Four Futures for AI: Possible Trajectories of AI Development to 2030. OECD AI Papers, February 2026. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2026/02/exploring-possible-ai-trajectories-through-2030_b6fb75d9/cb41117a-en.pdf
70. Stanford HAI. 2026. AI Index Report 2026. Op cit.

Figure 4. AI scenario implications for stewardship

CAPABILITY TRAJECTORY BY 2030

PROGRESS STALLS	PROGRESS SLOWS	PROGRESS CONTINUES	PROGRESS ACCELERATES
I capabilities remain broadly at current levels. Systems perform well-scoped tasks efficiently but require substantial human support. Diffusion and application development of existing capabilities continues.	Continued but slower gains. By 2030, AI handles well-scoped tasks taking humans hours to days. Improved agentic capabilities emerge in structured settings. Human oversight remains necessary for complex or consequential decisions.	Rapid continued progress in line with recent trends. By 2030, AI performs many professional tasks in digital environments equivalent to a month of human work. Operation becomes substantially more autonomous within human-set bounds. Agentic deployment spreads beyond specialist settings.	AI systems approach or match human capabilities across most cognitive dimensions. At the upper end of this scenario, which the OECD identifies as including AGI-level capability as a plausible variant, systems can pursue broad strategic goals autonomously and begin contributing to their own development.
OBSERVABLE SIGNALS			
Benchmark performance plateauing; task horizon doubling time extending markedly beyond current pace; slowing frontier model release cadence.	Task horizon doubling time extending but not stalling; continued organisational adoption growth; documented AI incidents rising with deployment volume.	Task horizon doubling continues across multiple domains; agentic deployment rising from current single-digit business function penetration; model documentation remaining patchy despite regulatory pressure.	Task horizon doubling time accelerating; rapid emergence of capable agentic systems beyond controlled enterprise settings; growing evidence of AI-assisted AI development at frontier scale; regulatory capacity visibly outpaced by deployment pace.
PRIMARY SOCIAL RISK CHANNEL			
The governance gap widens as adoption spreads without corresponding organisational adaptation. Deployment quality across the large population of firms already using AI becomes the dominant risk exposure.	A policy-to-practice gap widens across deployers as adoption continues to spread faster than governance adaptation. Workforce transformation pressures become more visible, including junior pipeline compression and wider algorithmic management.	Deployer dependence on upstream model and infrastructure providers intensifies. Formative work erosion accelerates. Human oversight thresholds become harder to sustain as system capability rises. The agentic governance gap widens across multiple sectors simultaneously.	At the lower end of this scenario: accelerated workforce disruption, concentration of economic and operational power in few AI providers, and deployer controls under severe pressure. At the upper end, governance question shifts from organisational controls to systemic architecture: the adequacy of oversight over AI-enabled AI development, and whether institutional capacity can keep up.
STEWARDSHIP EMPHASIS			
Whether firms are translating existing adoption into controlled, accountable use: use case inventory, oversight adequacy and remedy for currently deployed systems.	Whether governance is keeping pace with adoption; workforce impacts and formative work erosion; vendor governance as AI moves into more consequential processes.	Phased rollout and rollback controls; evidence that human oversight remains meaningful under operational pressure; developer-deployer accountability for remedy; infrastructure governance at scale.	At the lower end, phased rollout, rollback capability and infrastructure governance take on substantially greater urgency. At the upper end, firm-level engagement should combine with policy engagement and collective action. Deployer-side controls at the centre of this paper remain the main lever but require systemic complements at the upper end of this scenario.

Source: Amundi

Two observations emerge from this analysis. There is a threshold within the Progress Accelerates scenario at which firm-level governance, though still necessary, becomes insufficient as the primary response. Below that threshold, the deployer-side controls at the centre of this paper remain the main lever: use case discipline, meaningful human oversight, vendor accountability and grievance architecture. Above it, systems capable of operating with broad strategic autonomy and contributing to their own development create a governance problem that is systemic in character, extending beyond the reach of organisational controls. Investor stewardship remains relevant at that upper end, through policy engagement, collective action on assurance infrastructure, and scrutiny of developer-side controls on capability deployment, but the scope of the problem exceeds what firm-level engagement alone can address. The pace of capability progress is itself partly shaped by the social and institutional conditions this paper examines. Resource constraints on infrastructure expansion, public acceptability of data centre siting, regulatory friction on agentic deployment, and the availability of data labour on which training depends are all factors investors can scrutinise and on which they can engage. The scenarios are not purely exogenous constraints to which stewardship must respond: the conditions under which AI scales are partly determined by the governance and legitimacy questions at the centre of this paper.

VI. Investor engagement priorities

Below we set out key expectations on the social dimensions of AI governance, as a focused summary from the broader framework landscape reviewed above, organised around four governance functions that apply across the AI value chain. These expectations hold across all AI capability trajectories. Investors should continue to revisit the weight we place on these themes as the trajectory of AI development becomes clearer.

1. **Know your AI systems** (applies to: all companies; documentation and supply chain labour expectations weighted towards developers; human rights due diligence expectations applicable across the value chain)

- Maintain a current inventory of where AI is used, which uses carry higher risk and what role the company plays in the AI value chain, whether as developer, deployer or both.
- Document model capabilities, limitations and intended uses with sufficient specificity to support downstream governance and investor assessment.
- Maintain visibility over vendor and supply chain dependencies, including the labour conditions of workers involved in data annotation, content moderation and other AI supply chain functions.
- Conduct human rights due diligence across AI-related activities and business relationships, treating affected workers, users and communities as rights-holders rather than risk categories, and integrating findings into governance and deployment decisions rather than treating them as a standalone compliance exercise.

2. **Govern your AI systems** (applies to: all companies; responsible use policies weighted towards developers; vendor and supply chain governance weighted towards deployers but applicable to developers in relation to downstream distribution)

- Maintain governance arrangements with genuine authority: the capacity to delay or withdraw AI deployment where risks change or controls prove inadequate, not only to advise.
- Apply approval discipline before deployment in consequential settings and ensure human oversight is substantive rather than procedural, with reviewers having sufficient time, information and escalation routes to exercise real judgement under operational pressure.
- Govern third-party and vendor AI systems through structured due diligence, in-context testing and ongoing monitoring, with contractual accountability levers and escalation arrangements that extend through the deployment lifecycle rather than ending at procurement.
- Require vendors and downstream integrators to meet governance standards consistent with the company's own, and maintain visibility over how AI systems behave once deployed in third-party settings.
- For developers, maintain explicit responsible use policies governing military, law enforcement and surveillance applications, enforced through customer screening and contractual accountability.

3. **Account for AI systems' impacts** (applies to: all companies; deployment, consumer harm and disclosure expectations weighted towards deployers; documentation, downstream support and disclosure of model limitations weighted towards developers)

- Take active responsibility for AI's effects on the people it touches: workers, users, customers and communities.
- Where AI informs consequential decisions affecting access to services, employment or livelihoods, conduct workflow-level assessment in the actual deployment context, monitor for population-level patterns across demographic groups, and maintain workable routes to explanation, challenge and correction.
- Give workers credible internal routes to raise concerns about AI deployment, and ensure that where AI informs employment decisions, appeal and review routes function in practice.
- Where AI systems interact with or affect children or other vulnerable users, apply stronger safety design, accessible escalation and proactive harm monitoring.
- Disclose material AI uses, associated risks and governance arrangements with sufficient specificity to allow investors and affected stakeholders to assess whether accountability is genuine, moving beyond principles statements towards operational evidence of how risks are identified, managed and remediated.
- Ensure that responsibility for remedy and incident response is clearly allocated and functional across developer-deployer relationships.

4. **Govern adaptively** (applies to: all companies; workforce and formative work expectations weighted towards deployers; infrastructure and social licence expectations applicable to developers and deployers with material compute operations; vendor dependence expectations applicable across the value chain)

- Build governance that keeps pace with the speed of AI deployment and capability development.
- Monitor AI's effects on the character of work and developmental pathways, not only on headcount, and take active steps to ensure that productivity gains from AI are reinvested in workforce capability rather than allowing review capacity to thin as output expands.
- Treat social license as an ongoing governance question: for companies with material infrastructure operations, engage communities early in project development, allow local concerns to influence project design, and improve the specificity of disclosure on where resource pressures are concentrated and how trade-offs between expansion and community needs are resolved.
- Revisit governance arrangements as AI capability develops, with particular attention to the adequacy of controls as more autonomous systems enter consequential settings, and to the governance implications of increasing dependence on a small number of upstream model and infrastructure providers.

Amundi's approach to engagement on ethical AI

Amundi engages with companies on AI ethics directly and through collective groups, including the Ranking Digital Rights engagement convened by the Investor Alliance on Human Rights, the collective Big Tech and Human Rights engagement convened by the Council on Ethics of the Swedish AP Funds, and the World Benchmarking Alliance's Ethical AI engagement where Amundi is also a Steering Committee member.

Recognising the impact of technology and the growing importance of AI in all sectors, we also directly engage technology corporates where AI ethics issues have been identified as material, including companies involved in the development and deployment of AI (e.g., software and interactive media companies), as well as companies located further up in the AI value chain (e.g., semiconductors) who nonetheless need to meet their downstream human rights and ethics due diligence obligations in response regulatory and reputational risks associated most prominently with product misuse.

As well, we are engaging with companies outside of the technology sector, such as consumer goods and services, pharmaceutical, medical technology and medical device manufacturers with increasing exposure to AI who lack formal AI oversight and governance but nonetheless face increasing risks associated with the nature of their AI use. Details of our engagement efforts can be found in our Engagement Report[71].

71. Past and most recent engagement reports can be found at: https://www.amundi.com/institutional/responsible-investment-policies-and-reports

Disclaimer

Publication date: June 2026

Doc ID: 5567608

Document issued by Amundi Asset Management, "société par actions simplifiée" – SAS – Portfolio manager regulated by the AMF under number GP04000036 - Head office: 91-93 boulevard Pasteur – 75015 Paris – France – 437 574 452 RCS Paris – www.amundi.com – Photo credit: Getty Images, iStock – Setting-up: Atelier Art6.