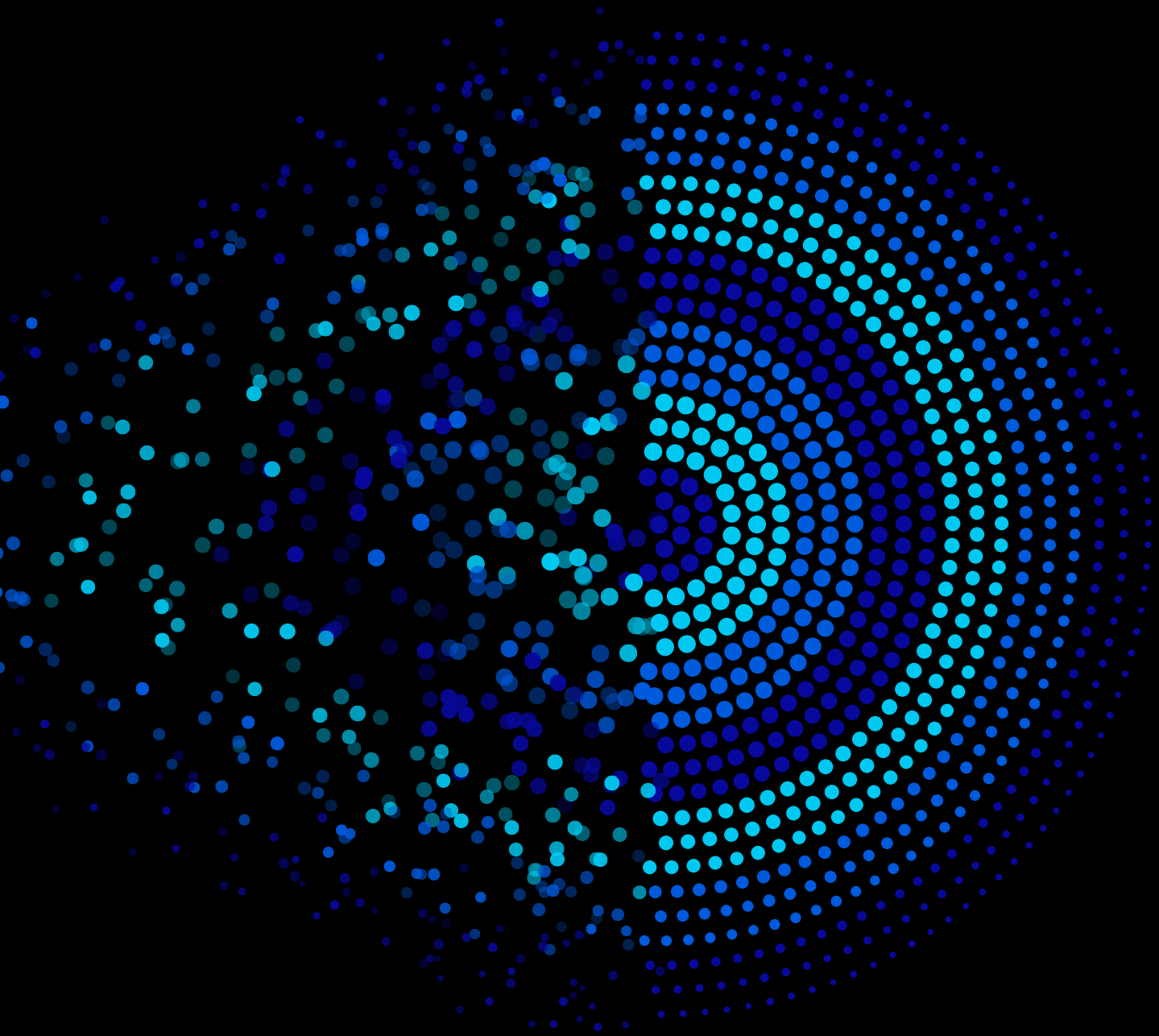


IoD BUSINESS PAPER

AI Governance in the Boardroom

The essential governance questions for your next board meeting

This paper was produced by the Institute of Directors Expert Advisory Group on Science, Innovation and Technology in collaboration with the IoD's Head of Innovation and Technology Policy.

With thanks to the members of the Expert Advisory Group:

Michael Ambjorn, CEO PropTech Denmark

Dr Phil Clare, CEO Queen Mary Innovation

Dr Paul Corcoran, Managing Director Interimconsult

Dr Jonathan R. Everhart, CEO & Chief Investment Officer Global ReEnergy Holdings

Faisal Khan, Chair Institute of Directors Expert Advisory Group on Science, Innovation and Technology

Rob Noble, Chair The Webinar Vet & Fidem Consulting

Pauline Norstrom, CEO Anekanta®AI and Anekanta®Consulting

Yinka Olarinde, Principal Consultant, Strategy & Technology, Siep Global Consulting Ltd

And with thanks to:

Dr Erin Young, Head of Innovation and Technology Policy, Institute of Directors

Contents

Introduction and Executive Summary

While Artificial Intelligence (AI)* in various forms has been in use for decades, developments in Generative AI in recent years have sparked widespread discussion of its potential benefits and risks across the economy and wider society. With responsibility for overseeing organisational strategy and risk management, directors should seek to harness the opportunities and mitigate the risks of AI, where relevant; this requires a robust and adaptable approach to governance.

In our first edition 'AI in the Boardroom', published in 2023, we identified 12 principles that boards should consider when implementing a governance framework for AI in their organisations. Capabilities and applications have advanced and scaled significantly since then, and AI tools, technologies and systems are now undergoing further massive investment by Big Tech, governments and other stakeholders in a so-called 'global race' for profit, control and geopolitical influence. AI is no longer only technical; it has become a strategic, operational, ethical, legal, and political issue, with many unknowns about the the mature state of products, ecosystems and value chains.

Through its AI Opportunities Action Plan, the UK Government has positioned AI as a pillar of its growth strategy. While the plan promotes innovation and rapid adoption to unlock economic value, we cannot lose sight of the commercial, regulatory, reputational and other sociotechnical risks and systemic harms that AI may propagate, both for organisations and society more broadly, including misinformation, security threats, IP infringement and disruption of employment.

The Institute of Directors is actively engaged with the UK Government's Department for Science, Innovation and Technology (DSIT) across tech policy priorities, including on the delivery of recommendations from this plan, such as supporting private sector AI adoption.

AI adoption and governance are highly context-specific. As the guiding questions and principles in this paper suggest, there is no`one-size-fits-all' approach. Effective governance depends on a range of organisational factors including sector, size and maturity, and must be adaptable to reflect the constantly evolving technology and policy landscape.

This paper is designed to complement data and AI strategy decisions by offering practical guidance for boards and directors of organisations using or planning to use AI systems, and should be seen as a resource for oversight rather than a comprehensive document. Despite limited capacity as they confront other urgent challenges, a deeper understanding of AI's operational and societal impacts is an essential boardroom priority. Boards must commit to agility, a culture of curiosity and innovation, and continuous learning in order to responsibly steer AI's strategic application in alignment with core organisational values and long-term goals.

*We recognise Artificial Intelligence (AI) is a contested, vague and 'catch-all' term. See Appendix A for a broad working definition for the purposes of this paper.

As boards explore these questions, they are encouraged to think beyond compliance and towards strategic foresight. Consider:

- What if AI systems rendered your organisation or business model obsolete, via disintermediation or other structural shifts? What new models or missions might emerge?
- What is the current AI maturity and readiness of your organisation? Are you clear on what and how data - and predictive, generative and agentic tools - are used across your operations and functions, including through third-party systems and across the supply chain and tech stack? Are they being used responsibly?
- What if the energy demands of AI conflict with your climate commitments? What are the creative trade-offs?
- What if power or access to infrastructure is lost for an extended period? What does resilience look like in this context?
- What is your risk appetite - and capital/ resource allocation - when it comes to strategic AI adoption, integration and development? What security measures, practices and protocols can you embed across the lifecycle?
- Have you clearly articulated any 'AI-shaped' problems aligned with long-term business objectives, strategic priorities and high-level values? How can success and failure be measured and evaluated?
- Does the board have appropriate skills and experience to support effective oversight? Are changes to board structure necessary? What roles should the management team and workforce play?
- Have you considered how to value the Return on Investment (ROI)? For example, whether your aspirations are to deploy AI for efficiency and productivity or for more radical business transformation, do your plans consider the benefits, impacts and risks holistically?

These actionable questions are designed to unlock innovative thinking around corporate governance, and challenge boards to anticipate disruptions before they arrive.

Boards are also encouraged to cross-reference their AI governance work with:

IoD Science, Innovation and Technology policy resources, including blogs, events and podcasts

IoD Director Competency Framework

IoD - London Business School Policy Paper 'Assessing the expected impact of generative AI on the UK competitive landscape'

IoD Code of Conduct for Directors: Leading by Example, Integrity, Transparency, Accountability, Fairness, Responsible Business

Search 'IoD Glossary of Science, Innovation and Technology' terms

Insights from the IoD Policy Voice survey on AI adoption

The March 2025 IoD Policy Voice survey, capturing the views of nearly 700 directors and business leaders (across sectors, business sizes and regions) reveals a dynamic yet cautious landscape around AI adoption. Findings underscore the critical need for board-level clarity, strategic oversight and skills development in the adoption and governance of AI technologies and systems.*

Key themes

1. Increasing adoption, but governance gaps remain

- Nearly two-thirds of directors now personally use AI tools to aid their work.
- Half of directors report that their organisation uses AI across any of its functions and processes.
- Despite growing experimentation, a quarter are concerned about the lack of an internal AI policy, strategy or data governance framework in their organisation.

2. Benefits are recognised, but scepticism persists

- Increased productivity, operational and administrative efficiencies across functions, and better data insights and analytics are the top cited benefits.
- However, many directors remain sceptical about AI's business value, citing overhyped claims and lack of accuracy as significant concerns.
- There's a notable tension between enthusiasm for efficiency gains and concerns about reliability and implementation, with the latter expressing the need for tighter governance before adoption.

3. Skills gaps, lack of trust and security risks are the biggest concerns

The biggest barriers to AI adoption include:

- Limited expertise or understanding of models and tools at management and board level
- Lack of trust in AI outcomes (e.g. explainability, reliability, accuracy)
- Security risks (e.g. cyber, data protection and privacy)
- Skills, training and knowledge gaps at all organisational levels
- Safety and ethical risks, and societal impacts (e.g. bias, fairness)

* For the purposes of this survey, respondents were asked to consider 'AI technologies' broadly, across the full range of current capabilities (including generative, predictive and agentic).

A set of reflective checklists

The checklists that follow are designed to help you establish a board-level understanding of your organisation's position on AI. It draws on a set of 12 principles first developed by Pauline Norstrom of Anekanta®Consulting in 2020, which can help guide the responsible use of AI throughout an organisation.

The 12 principles have been revised for 2025 to integrate new legislation, standards (including ISO/IEC 42001 and 5259), best practices, and boardroom realities, retaining the original structure and practical tone that made the first edition accessible and impactful.

It is clear that AI must now firmly be on the board agenda, and considered as an essential part of governance responsibilities. AI should not be confined to the domain of the IT function - although the Chief Information Officer (CIO) / Director of IT may be responsible for its implementation and management. Some organisations may also have Chief Information Security Officers (CISOs), tasked in line with the title. Nevertheless, board oversight and organisational governance are essential.

Updated 12 principles

1. **Monitor the evolving regulatory and (geo)political environment**
2. **Continually audit and measure** what AI is in use, along with principles, processes and controls
3. **Undertake impact and risk assessments** that consider the business and its wider stakeholder community
4. **Establish board accountability and management responsibilities** for AI governance
5. **Set high-level strategic goals** for AI adoption aligned with the organisation's values and business objectives
6. **Empower a cross-functional, operational independent review committee**
7. **Validate, document and secure data sources, and assess data assets**
8. **Train and upskill people** to use AI effectively and responsibly, and embed in the culture
9. **Comply with privacy requirements,** including those set out in relevant data protection legislation
10. **Comply with security-by-design requirements** to ensure systems are cyber resilient
11. **Test and evaluate systems and remove from use** if unintended impacts or harms are discovered
12. **Review systems, policies and governance practices regularly**

Monitor the evolving regulatory and (geo)political environment

For board directors, AI regulation is no longer a speculative concern; it is a growing strategic, legal, and ethical issue. Boards must understand how national policy, international standards, and foreign legislation may impact their business operations, directly or indirectly.

United Kingdom: sector-led regulation

As of June 2025, the UK Government is taking a decentralised approach to AI regulation, with no central legislation. Instead, governance has been delegated to existing sector-specific regulators such as the Financial Conduct Authority (FCA), Information Commissioner's Office (ICO), Competition and Markets Authority (CMA), Equality and Human Rights Commission (EHRC), Office of Communications (Ofcom) and the Medicines and Healthcare products Regulatory Agency (MHRA).

The Digital Regulation Cooperation Forum (DRCF), established in 2020, brings together four of these UK regulators with responsibilities for digital regulation (CMA, FCA, ICO and Ofcom). DRCF members work together to support AI regulation, particularly for the most advanced models, in a way that both promotes benefits, and mitigates the risks to people and competition.

The UK's implementation model is coordinated across regulators, and supported by departments and institutions, including the UK AI Security Institute, a directorate of DSIT. The AI Opportunities Action Plan underscores this approach, with recommendations urging government to 'commit to funding regulators to scale up their AI capabilities' (Recommendation 25), and to 'work with regulators to accelerate AI in priority sectors and implement pro-innovation initiatives like regulatory sandboxes. These should be targeted in areas with regulatory challenges but high-growth potential' (Recommendation 27).

Read more:

AI Opportunities Action Plan (2025)

AI Cyber Security Code of Practice

'A pro-Innovation approach to AI regulation' white paper (2023)*

Digital Regulation Cooperation Forum (DRCF)

AI Security Institute

Artificial Intelligence (Regulation) Bill

Data (Use and Access) Act 2025

*Five cross-sector principles, derived from the OECD AI Principles, were outlined in this paper, including: safety, security, and robustness; appropriate transparency and explainability; fairness; accountability and governance; and contestability and redress.

European Union: The AI Act and Product Liability Directive

For UK-based organisations operating in the EU, or working with EU-based partners, the AI Act represents a fundamental shift in regulatory expectation:

- Entered into force in August 2024
- Introduces a risk-based framework (unacceptable, high, limited, minimal risk)
- Places the OECD-originated High-level Experts Group AI Principles on a statutory footing for high-risk systems
- Imposes obligations on providers, deployers, importers, and distributors of AI systems
- Places AI Literacy on a statutory footing for all AI systems regardless of risk level, requiring appropriate training on the use of AI by those interacting with it or using its decisions.

High-risk systems must meet mandatory requirements around:

- Human oversight
- Data governance
- Technical robustness and accuracy
- Transparency, technical documentation and traceability
- CE marking, conformity assessments and post-market surveillance

Alongside the AI Act, the amended Product Liability Directive which is in application from December 2026, modernises EU product liability law, enabling compensation claims for defective AI products under national liability rules. UK organisations providing AI services in the EU will be subject to both regimes.

What boards should consider

- [ ] Are we clear on the **regulatory expectations** for AI use within our sector?
- [ ] Have we **mapped our exposure** to both UK and EU regulatory frameworks (and others where relevant)?
- [ ] Are we aligned with **non-statutory standards** such as ISO/IEC 42001 or the UK AI Cybersecurity Code of Practice?
- [ ] Are we ensuring that monitoring of the regulatory and (geo)political environment is embedded into the R&D initiatives of the organisation?
- [ ] Are we using **risk-based tools** to evaluate and monitor AI systems throughout their lifecycle?
- [ ] Are we engaging with **regulators, industry forums**, or **sandbox initiatives** (e.g. ICO or FCA) to stay ahead of evolving expectations?

Read more:

EU AI Act

Product Liability Directive

Tools such as the Global AI Regulation Tracker can help boards monitor the evolving regulatory environment globally: https://www.techieray.com/GlobalAIRegulationTracker

Continually audit and measure what AI is in use, along with principles, processes and controls

To govern AI responsibly, directors must ensure that all AI systems in use across the organisation - whether developed in-house or sourced from third parties - are identifiable, auditable, and measurable. This must include acknowledgment that 'shadow' AI use - where employees independently use AI tools such as LLMs to assist with their work - is common practice. The organisation's principles for AI use must be embedded in formal governance systems and integrated into operational management. For many, this will mean aligning with recognised quality and risk standards such as:

ISO 9001:2015 (Quality management systems - Requirements)
ISO 9001 is a globally recognised standard for quality management. It helps organisations of all sizes and sectors to improve their performance, meet customer expectations and demonstrate their commitment to quality. Its requirements define how to establish, implement, maintain, and continually improve a quality management system (QMS).

ISO/IEC 42001:2023 (Information technology - Artificial intelligence - Management system)
ISO/IEC 42001 is an international standard that specifies requirements for establishing, implementing, maintaining, and continually improving an Artificial Intelligence Management System (AIMS) within organisations. It is designed for entities providing or utilising AI-based products or services, ensuring responsible development and use of AI systems.

ISO/IEC 27001:2022 (Information security, cybersecurity and privacy protection - Information security management systems - Requirements)
ISO/IEC 27001 is the world's best-known standard for information security management systems (ISMS). It defines requirements an ISMS must meet. The standard provides companies of any size and from all sectors of activity with guidance for establishing, implementing, maintaining and continually improving an information security management system.

NIST AI Risk Management Framework (AI RMF) In collaboration with the private and public sectors, NIST has developed a framework to better manage risks to individuals, organisations, and society associated with AI. The NIST AI Risk Management Framework (AI RMF) is intended for voluntary use and to improve the ability to incorporate trustworthiness considerations into the design, development, use, and evaluation of AI products, services, and systems.

AI Standards Hub
The Hub's mission is to advance trustworthy and responsible AI with a focus on the role that standards can play as governance tools and innovation mechanisms. It aims to help stakeholders navigate and actively participate in international AI standardisation efforts and to inform the direction of these efforts.

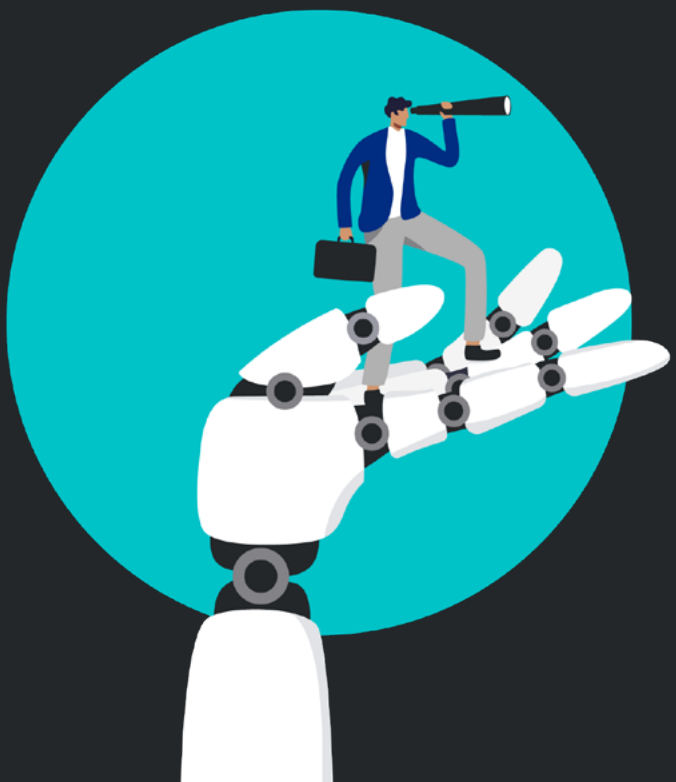
*See also: ISO/IEC TR 24030:2024 for AI use cases across domains

While ISO 9001:2015 provides a foundation, ISO/IEC 42001 offers a more tailored management system for AI, designed to be adopted across the lifecycle of AI systems. It includes guidance on:

- Setting an acceptable use AI policy governed at board level
- AI risk and impact management and accountability
- Performance evaluation and continuous improvement
- Documentation and transparency
- Embedding ethical principles into governance processes

Boards must also recognise that auditing AI is not a one-off task. It must be a continuous and evolving process, as systems learn, adapt, or integrate new datasets. Without routine audits, risks may go undetected, and systems may diverge from their intended purpose.

Furthermore, organisations intending to report on AI in their annual report or ESG disclosures will need to define internal mechanisms for evaluating effectiveness, fairness, and alignment with business objectives.



What boards should consider

- ☐ What is our **risk appetite** when it comes to AI use? Are AI-related risks explicitly captured in our **Risk Register**?
- ☐ Do we have an **AI Policy**? Does it cover the organisation as well as the supply chain?
- ☐ Does our organisation have a well-communicated **responsible AI framework**? How was it developed, and **what principles does it prioritise**? How is it **assessed in practice**?
- ☐ Do we have a real-time **inventory of AI systems** in use - including those embedded in third-party and supplier tools across the stack?
- ☐ Does the **audit committee** (or its equivalent) have oversight of AI systems?
- ☐ Are our **ethical principles clearly articulated**?
- ☐ Have we ensured that these principles are also **machine-readable**, enabling technical teams to embed them into code and systems architecture?
- ☐ Are our AI governance standards embedded into our **quality management systems**, such as ISO standards?
- ☐ Do we plan to report on AI usage, impact or risk mitigation in our **annual report** or **ESG reporting**?
- ☐ If yes, have we set an appropriate timeline for this? Would external **advisory or assurance support** be helpful?
- ☐ How do our **existing privacy, cybersecurity and data governance policies address AI**, if at all?

Spotlight on technical infrastructure integration: machine-readable formats and the foundations of digital efficiency*

As organisations increasingly rely on data-driven systems, the ability to process information rapidly and accurately becomes mission-critical. Machine-readable formats - data structures specifically designed for automated interpretation - underpin everything from regulatory compliance and financial reporting to real-time decision support and AI integration.

When organisations store or share information in human-readable-only formats, it limits automation, increases the risk of manual errors, and creates bottlenecks. By contrast, properly structured machine-readable data enables seamless integration across platforms, faster insights, and more resilient digital operations.

Examples:

- A scanned PDF of a printed table is human-readable - people can understand it by looking at it, but software cannot easily extract the numbers without error-prone OCR techniques.
- A CSV file containing the same table is machine-readable - computers can instantly parse, analyse, and connect the data with other systems without human intervention.
- A modern passport represents a hybrid - it contains human-readable information printed visibly for border agents, machine-readable text zones for optical scanners, and often an embedded RFID chip that stores structured digital data for secure automated verification.

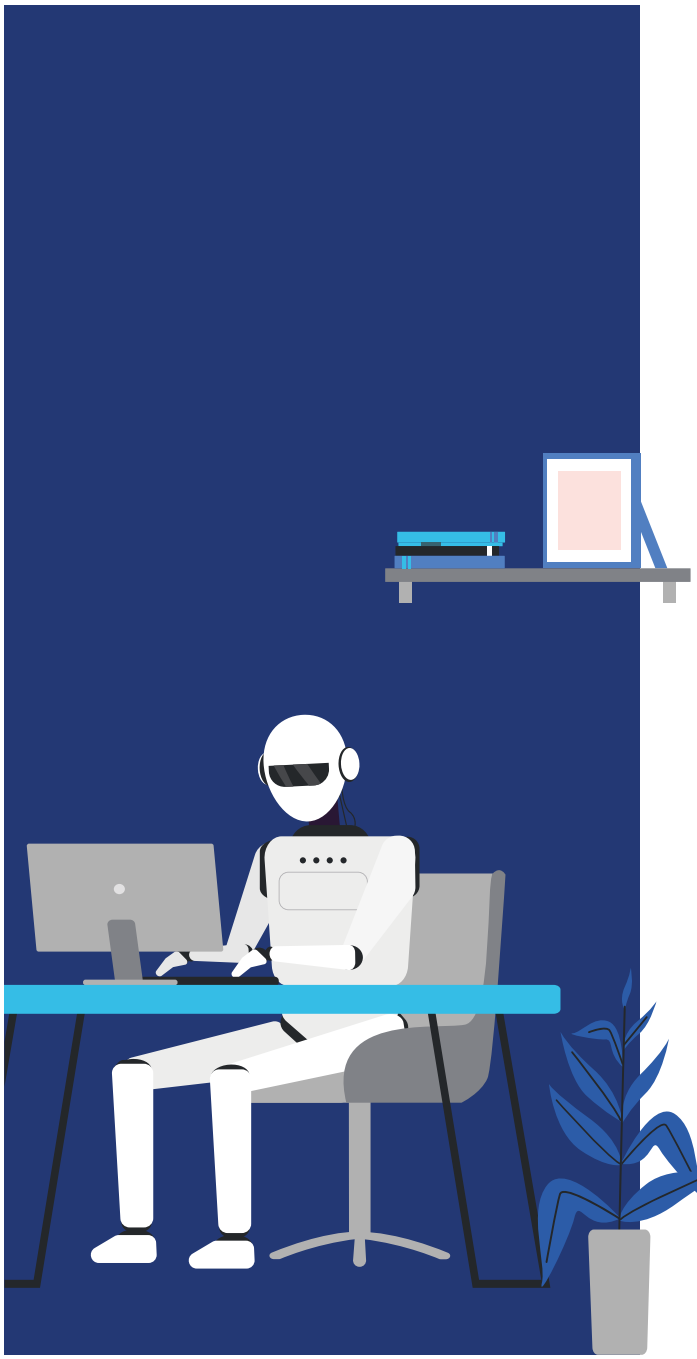
This issue is becoming even more significant as governments and regulators worldwide begin to mandate machine-readable disclosures in areas like environmental, social, and governance (ESG) reporting, financial filings, and public-sector transparency initiatives. Organisations that fail to adapt may find themselves unable to meet evolving compliance demands - or unable to leverage emerging AI capabilities effectively.

What boards should consider

- [ ] Have we, as an organisation, mapped where critical internal or external information exists only in human-readable formats (e.g. PDFs, presentations, unstructured reports)?
- [ ] Are we investing in systems and standards that prioritise structured, **machine-readable outputs** across key functions?
- [ ] When procuring new technology platforms, do we assess the machine-readability and **interoperability** of their data?
- [ ] Are leadership teams aware of how machine-readable data underpins AI readiness, automation, and future regulatory compliance?

*See also: Validate, document and secure data sources, and assess data assets

Undertake impact and risk assessments that consider the business and its wider stakeholder community



Impact assessments are a foundational element of responsible AI governance. They help boards and executive teams understand the implications of AI use - not just in technical terms, but in relation to people, policy, culture and purpose.

Risk management that accelerates rather than blocks innovation is critical. Boards should ensure that comprehensive AI impact assessments, including risk assessments, are carried out across the organisation, with particular focus on:

Workforce Impact

Assess whether:

- Roles may be augmented, reshaped or displaced by AI
- Tasks are being automated in ways that reduce or shift human responsibility
- There is sufficient transparency, training, and engagement for those affected
- Changes comply with employment, privacy, health & safety, and equality laws

Employees should be:

- Clearly informed when they are interacting with or affected by AI (e.g. an HR chatbot)
- Given the opportunity to provide feedback

Broader stakeholder impact

Impact assessments should also consider:

- Customers (e.g. changes to service experience, personalisation, fairness)
- Suppliers and partners (e.g. contractual or operational dependencies on AI and supply chain resilience)
- Shareholders and investors (e.g. risk appetite, brand and reputation implications)
- Regulators, communities and society (e.g. data use, misinformation, or systemic bias)

In high-risk contexts, organisations should consider commissioning independent third-party assessments or leveraging AI-specific tools such as:

- Algorithmic impact assessments
- Bias audits
- Data protection impact assessments (DPIAs) (as required under DPA 2018/UK GDPR)
- Equality impact assessments (EqIAs)
- ISO/IEC 42005 (Information technology — Artificial intelligence — AI system impact assessment)

Explainability must also be considered:

- Are the decisions or outputs of the AI system understandable to the people they affect?
- Can humans intervene or override outcomes if needed?

What boards should consider

- [ ] Is all AI use **clearly and appropriately** labelled across the organisation, including systems with which employees and customers interact?
- [ ] Do we have a **structured, repeatable process** for evaluating the impact of AI on stakeholders, including one that enables **external input**?
- [ ] Are our AI-enabled decisions **sufficiently explainable** such that individuals impacted by the outcome can understand how the decision was made?
- [ ] Do we **understand the impact of AI** on our supply chain, their resilience, and business models?
- [ ] Is stakeholder feedback **actively invited and responded to**?
- [ ] Where appropriate, have we considered **independent assurance or third-party assessment**? Similarly, impact assessments must be undertaken for all stakeholder groups including customers, suppliers and partners.

Governance is not only about control, but about engagement. Impact assessments help organisations build trust and resilience, and prepare for the multifaceted demands of scale.

Establish board accountability and management responsibilities

AI cannot (and must not) be considered solely the responsibility of technologists or operational leaders. Its adoption and governance have strategic, operational, ethical, reputational, and legal implications. As such, accountability must reside at board and management level.

Directors are responsible for overseeing how AI is used throughout the organisation, including systems with which employees and customers interact:

- Proprietary systems
- Tools sourced from third-party vendors, across the tech stack
- Embedded AI within platforms and services

The board must also ensure that AI use:

- Aligns with the organisation's values
- Is transparent and auditable
- Does not create discriminatory, unsafe, or disproportionate outcomes
- Has a clear governance structure, with named accountability for oversight

The board should retain a final veto over the implementation or continued use of AI systems, particularly where commercial, reputational, safety or regulatory risks are identified.

AI governance should be:

- Embedded in board-level risk frameworks
- Reflected in terms of reference for board committees (e.g. audit, risk, ESG, remuneration)
- Supported by dashboards, assurance reports, and regular updates from the executive and independent review committee

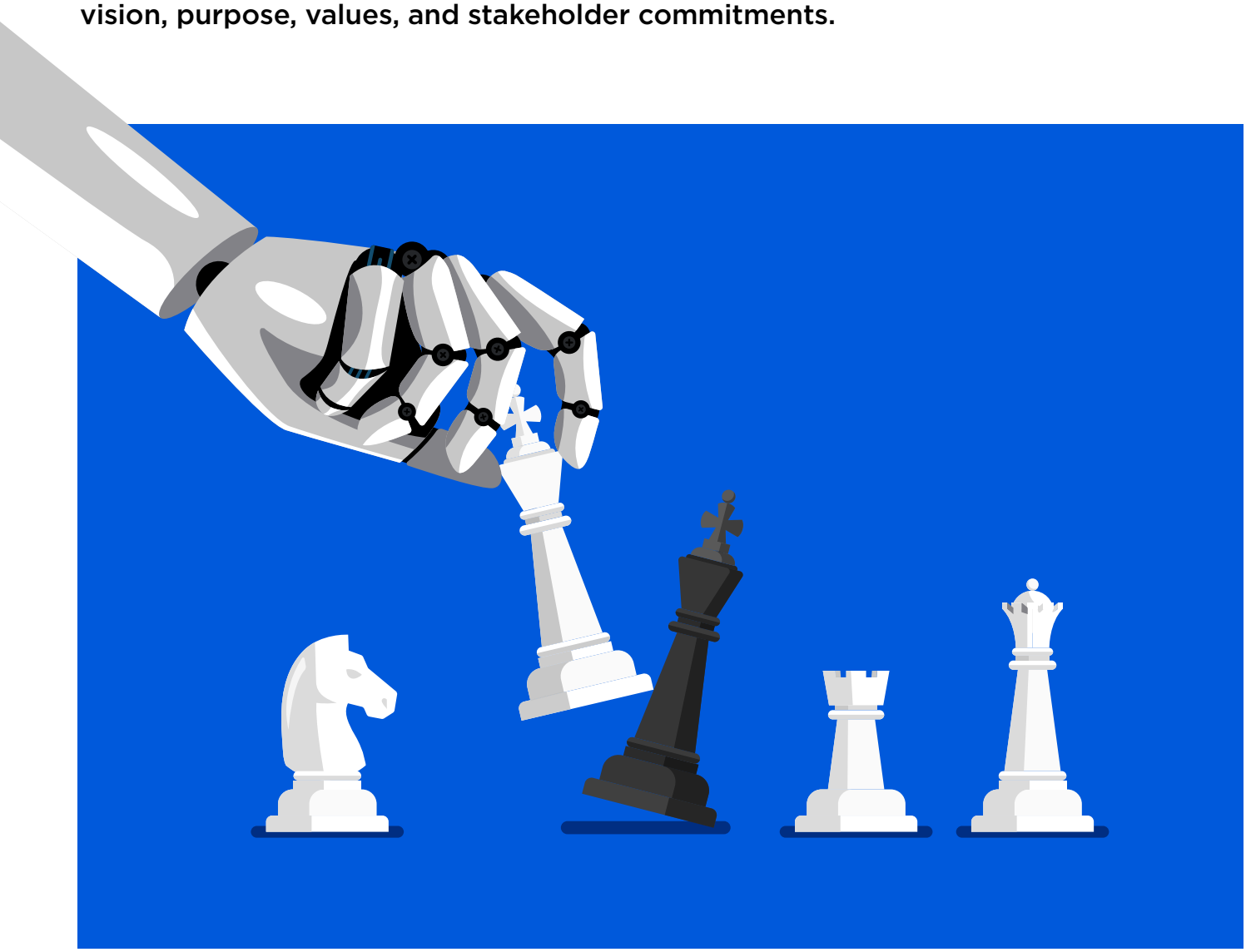
What boards should consider

- [ ] Does the board have the **capability** and **confidence** to evaluate AI-related risks and opportunities?
- [ ] Are we clear on how **data, generative and predictive tools** are used across our operations including through third-party systems across the tech stack?
- [ ] Has the board **formally identified a director or committee** with accountability for AI oversight? How about on the management team?
- [ ] How are we embedding **digital ethics** in our board discussions, strategy reviews, and committee structures?
- [ ] Do we actively communicate to stakeholders - including staff and investors - that AI is being used **responsibly**?
- [ ] If AI is already in use, are we confident we know **where, how, and why**?
- [ ] If AI is not yet in use, are we confident we understand where it may be **indirectly influencing our decisions** (e.g. via suppliers, partners, or data feeds)?
- [ ] Who is responsible for AI **system procurement**?
- [ ] Which individual or body at the board (or management) level is responsible for **data governance**?

Board responsibility should transcend legal compliance and embrace responsibility with AI governance, safeguarding stakeholder interests and building trust, while also meeting the fiduciary duty to their shareholders.

Set high-level strategic goals for AI use aligned with the organisation's values and business objectives

AI should not be introduced simply because it is novel or commercially available. Instead, every application of AI within an organisation should be guided by a clear, high-level set of goals, shaped by the organisation's vision, purpose, values, and stakeholder commitments.



Boards have a duty to ensure that AI use:

- Is strategic, not reactive
- Enhances value creation for customers, employees, investors, and wider society
- Reflects the organisation's ethical foundations and long-term direction

High-level goals help direct AI investments and use cases toward positive, measurable outcomes and can be used to evaluate systems during audits, ethical reviews, and impact assessments.

Examples of high-level goals might include:

- Augmenting human intelligence and creativity, rather than replacing it
- Improving the speed, consistency, and quality of decision-making
- Enhancing accessibility, inclusion, and fairness in products or services
- Protecting stakeholder wellbeing, ensuring no harm to employees, customers or communities
- Supporting climate and sustainability targets, including responsible use of energy and compute resources
- Driving innovation that aligns with the organisation's future capabilities and societal role
- Fostering a culture that openly encourages dialogue and understanding of the risks as well as the opportunities with AI for the organisation

AI goals can and should vary by function or department, but the strategic framing should be clear and consistent at the board level.

What boards should consider

- [ ] Have we clearly articulated what are our **'AI-shaped' problems, and what we would want AI to achieve** across the organisation?
- [ ] Do we understand how AI helps us to differentiate and bring **competitive advantage**? Are we leading the market or following the first mover?
- [ ] Are these goals written in **plain language**, understandable from the board to the frontline?
- [ ] Where applicable, are these goals **machine-readable**, so they can be embedded into AI models or operational code?
- [ ] Are the goals **aligned with the organisation's vision, purpose, values, and ESG commitments**?
- [ ] Are the **expected benefits of AI measurable** and have we defined what success or failure looks like?
- [ ] Have we considered **both near-term and long-term outcomes**, and how AI may evolve or adapt over time?

Well-crafted AI goals not only guide implementation, but also serve as touchstones for trust, giving stakeholders confidence that technology is being deployed deliberately, responsibly, and with purpose.

Empower a cross-functional, operational independent review committee

AI governance must be more than policy on paper; it should be actively practiced. At the heart of this operational governance may be a well-structured independent review committee to oversee AI, digital and/or data, equipped with the skills, mandate, and independence to make principled decisions.*

The board should empower this committee to:

- Review AI proposals, projects, and deployments at key stages
- Request or conduct risk and impact assessments, including transparency, bias, security, privacy, and societal considerations
- Provide guidance on alignment with organisational values and legal frameworks
- Serve as a channel for employee and stakeholder input on AI issues
- Raise red flags and recommend halts or adjustments when ethical risks are high

Critically, this committee should have the formal authority to pause, recommend revision, or veto AI projects that fail to meet agreed ethical, legal or operational standards.

To be effective, the committee must be:

- Diverse in both identity and discipline including representatives from HR, legal, IT, strategy, operations, and employee groups
- Embedded in governance processes, with a clear line of accountability to the board
- Properly resourced, trained, and supported
- Guided by a clearly articulated Terms of Reference and aligned with broader values of the organisation
- Proportionate to the size of organisation e.g. micro or SMEs may consider fractional or outsourced ethical oversight

The committee should not be a rubber stamp. It should be a deliberative, inclusive, and critically engaged forum, supporting both innovation and accountability.

*Resources including online courses can be very helpful here.

What boards should consider

- [ ] Does the independent review committee have a **clear and formal Terms of Reference**, approved by the board?
- [ ] Is the committee made up of a **diverse group of individuals** across roles, backgrounds, and perspectives?
- [ ] Have all members received **adequate training in responsible AI** and regulatory awareness?
- [ ] Is the **committee sufficiently empowered** to **veto**, **delay or reframe** AI projects when warranted?
- [ ] Are there formal processes for **staff or stakeholders to raise concerns** to the committee?
- [ ] How often does the committee report to the board and is this built into **regular governance cycles**?
- [ ] What **incident-reporting mechanisms are in place**?

Validate, document and secure data sources, and assess data assets

Data is the foundation of every AI system, and boards must ensure that its provenance, integrity and relevance are properly governed.

Before deploying AI, organisations must:

- Define the purpose and scope of the system
- Identify, assess, and document the sources of data used to train, fine-tune, or operate the system
- Evaluate the quality, completeness, and representativeness of the data, where possible
- Detect and mitigate bias, duplication, drift, or contamination, including risks associated with the growing use of synthetic data
- Consider the role of anonymisation, pseudonymisation and federation to prevent re-identification and inference which may impact on privacy rights

If data issues are discovered post-deployment, such as evidence of bias, exclusion, or model failure, organisations must be able to:

- Trace the root cause (via audit trails, version control, logs)
- Take corrective action, including retraining or suspension
- Document all adjustments to support transparency and accountability

The use of synthetic or AI-generated data must be explicitly declared and reviewed. AI systems trained on such data may carry self-reinforcing distortions if unchecked - a risk that compounds over time. Boards should ensure that teams adopt frameworks such as:

- ISO/IEC 5259 (AI data quality management bundle) - including AI data lifecycle, traceability and quality measurement
- FAIR principles[1] (Findable, Accessible, Interoperable, Reusable) - for metadata management
- AI Ethics and Governance in Practice: Responsible Data Stewardship in Practice
- Secure data governance practices in line with cybersecurity and privacy policies

Decision logic - whether rule-based, statistical, or AI-driven - should be:

- Understandable to humans
- Interpretable by board-level stakeholders, particularly in high-impact use cases
- Monitored with tripwires, logging and escalation protocols

1 The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016).

What boards should consider

- [ ] Is the decision logic behind our AI systems explainable to the board and relevant non-technical leaders? Or are we in 'black box' territory, with unquantified risk?
- [ ] Are tripwires and monitoring tools in place to alert us to nefarious capability, data or model drift?
- [ ] Are data logs secured, retained, and auditable over time?
- [ ] What **level of quality is our data**, and do we own it?
- [ ] Are the **sources of data clearly documented,** including whether any synthetic data was used?
- [ ] Is safety and ethics part of the sign-off process when new data sets or models are procured or created?
- [ ] Have we implemented Key Performance Indicators (KPIs) or quality metrics to track data integrity, bias, and lifecycle changes?
- [ ] Are the **issues around AI, copyright and the protection of our existing Intellectual Property (IP) understood**? How should we protect our IP and safeguard innovation?

By treating data as a strategic asset with ethical weight, boards can ensure that AI systems are trustworthy, defensible, and aligned with stakeholder expectations.

Spotlight: information provenance and the role of critical thinking

As AI systems increasingly generate, summarise, and transform information, organisations must take steps to ensure data provenance and content integrity. When AI is trained or prompted using unverified sources - or generates content without a clear basis - the risk of misinformation, reputational damage, or poor decision-making rises sharply.

This challenge is exacerbated by real-world developments. In 2024, Meta (Facebook's parent company) publicly withdrew from working with third-party fact-checking providers, signalling a broader shift away from structured content verification in some parts of the tech ecosystem. Boards must not assume that the information flowing into or out of AI tools has been independently checked.

This makes critical thinking and academic-style source evaluation vital across all layers of an organisation, particularly at the board level, where key decisions are often based on summaries, dashboards, and briefings.

What boards should consider

- [ ] Have we asked: "Where does this information come from?"
- [ ] Are AI-generated outputs, datasets, or visualisations being used in reports without **clear attribution or review**?
- [ ] Do we have a record of the **information sources and datasets used by our third-party vendors or suppliers**?
- [ ] Is the organisation investing in **data literacy and critical thinking skills**, not just technical upskilling?
- [ ] Are senior leaders trained to challenge assumptions and ask for sources?

Train and upskill people to use AI effectively and responsibly, and embed in the culture

For AI systems to be effective, fair, and aligned with organisational purpose and business objectives, it is not enough to focus on algorithms and infrastructure. People need to be empowered to use, question, and improve them.

Training, upskilling and awareness programmes must be designed to:

- Equip staff with practical understanding of the systems they are using in the enterprise, including how the field and models are evolving more broadly
- Build awareness of bias, fairness, and explainability and when humans must be in the loop
- Ensure responsible and transparent use of AI in day-to-day decisions
- Support foundational literacy and a culture of curiosity and innovation, not just technical competence

Training should be:

- Audience-specific: Tailored for frontline users, supervisors, technical teams, executives, decision-makers and other stakeholders
- Accessible and inclusive: Designed to engage people across roles and learning styles
- Continuous: Refreshed regularly to account for evolving systems, risks, and regulations
- Part of onboarding: New employees should be introduced early to the organisation's AI policies, tools, values, and expectations

Boards should view AI training not simply as a compliance task, but as an opportunity to:

- Build a resilient, informed workforce
- Enhance innovation readiness
- Prevent misuse of or blind trust in AI outputs

In sectors using generative or predictive AI, users must also be taught:

- How to verify AI outputs
- How to avoid over-reliance
- When to escalate concerns or intervene manually

What boards should consider

- [ ] Have we designed and delivered accessible, **audience-appropriate training** for all relevant teams, based on assessment of needs? Is this part of a broader change management programme?
- [ ] Are **frontline users** equipped with the tools and time to question or challenge AI outputs?
- [ ] Do staff understand the organisation's **ethical, safety or responsibility principles**, and how they apply to AI in practice?
- [ ] Are **new joiners onboarded into our AI governance expectations** as part of their induction?
- [ ] Do employees understand how **AI can reinforce bias**, and how to spot early warning signs?
- [ ] Have we created **incentives or recognition** for employees who build AI skills or contribute to governance?
- [ ] What consultation or communication is taking place with our workforce on potential AI impacts?

Boards that invest in widespread AI understanding are better placed to build trust, surface issues early, and ensure that human judgement remains a core part of decision-making.

Spotlight: AI accuracy and reliability as ongoing governance concerns

AI 'hallucinations', although a contested term, is often used to describe when an AI system (particularly a generative one) produces plausible but false or fabricated information. These outputs may seem coherent and credible, yet lack factual grounding.

This poses significant risks to organisations when:

- Inaccurate information makes its way into internal reports, briefings, or board packs
- Outputs are used for decision-making in areas such as compliance, communications, investor relations, or customer support
- Synthetic content is treated as verified knowledge

What boards should watch for:

- Overly confident language in AI outputs without a clear source or citation
- Sudden changes in tone, interpretation, or claims within reports or dashboards
- Use of AI-generated summaries, insights, or recommendations that lack traceability to verifiable data
- Absence of human verification or fact-checking processes, particularly for public-facing or strategic content

What boards should consider:

- ☐ Are we confident that outputs from generative AI (GenAI) tools are being **verified before use** in business-critical contexts?
- ☐ Have we established **guidelines or controls for AI-generated content,** particularly in regulated or reputationally sensitive functions (e.g. HR, Legal, Comms)?
- ☐ Are employees trained to **recognise hallucinated outputs**, and are boards aware of their potential presence in the materials they receive?
- ☐ Are the employees **prompting GenAI** sufficiently knowledgeable, competent and experienced in the subject matter? Are they professionally qualified and equipped to challenge hallucinations?
- ☐ Are we using **Retrieval-Augmented Generation (RAG)** techniques to ground outputs in trusted, up-to-date internal or external sources? If so, is the quality and provenance of the retrieved information being monitored?

Comply with privacy requirements

Privacy must be a core design principle, not a retrospective patch. Boards must ensure that AI systems are adopted, developed and deployed in compliance with relevant data protection laws and organisational policies.

In the UK and EU, this includes adherence to the principles of:

- GDPR (including lawfulness, fairness, transparency, data minimisation, and accountability)
- The Data Protection Act 2018 and UK GDPR
- Sector-specific guidance from regulators such as the Information Commissioner's Office (ICO); for example, see Guidance on AI and data protection, Explaining decisions made with AI, and the AI and data protection risk toolkit

Privacy-by-design means:

- Embedding privacy controls into the architecture and logic of the system
- Minimising the use of personally identifiable information (PII) where not strictly necessary
- Ensuring valid consent, data subject rights, and appropriate safeguards
- Building systems that are transparent and explainable, especially where decisions may affect individuals

AI development and engineering teams (if relevant) must be:

- Trained in the organisation's safety and ethics framework, where appropriate
- Equipped to assess when human involvement is required in decision-making
- Responsible for ensuring transparency in how data is used and protected
- Accountable to the independent review committee or equivalent governance structure

Boards must ensure that reporting lines are clear and that employees at all levels are confident in raising concerns around privacy or inappropriate use of data.

What boards should consider

- [ ] Are all data-driven and AI systems assessed for **privacy risks** before development or deployment?
- [ ] Do technical teams receive training on privacy, ethics, and regulatory obligations?
- [ ] Are employees able to **raise privacy-related concerns** via trusted, anonymous channels?
- [ ] Is the **reporting process understood across the organisation**, not just by technical or compliance teams?
- [ ] Are we confident that AI systems do not process **excessive or unnecessary personal data**?
- [ ] What **degree of understanding do our suppliers have of AI system privacy risks** and do we understand the implications?

A strong culture of privacy does more than protect the organisation from legal and reputational risk; it helps build trust with customers, staff, and regulators, and reinforces the board's commitment to fairness, responsibility and accountability.

Comply with security-by-design requirements

AI systems are only as trustworthy as they are secure. Boards must ensure that AI is designed, developed, and deployed with robust security controls in place, and that these are regularly reviewed and updated to respond to emerging threats.

Secure-by-design means that security is:

- Embedded from the outset, not added after deployment
- Built into data collection, model training, integration, and deployment processes
- Proportionate to the sensitivity and risk level of the AI's function
- Reviewed against recognised standards and certifications

Organisations may adopt recognised practices such as:

- Penetration testing, red teaming, and ethical hacking
- Secure software development lifecycles (SSDLC) for AI
- Encryption and access controls around training data, model outputs, and logs
- Review against frameworks such as Cyber Essentials Plus, ISO/IEC 27001, and the UK AI Cybersecurity Code of Practice (2025)

AI models, including LLMs and other multimodal models - especially frontier models - can evolve in ways that introduce new vulnerabilities. Boards must ensure they have visibility over:

- Reconfiguration risks
- Training data contamination
- Adversarial threats or attacks
- Risks of model inversion or data leakage

Just as financial systems are stress-tested, AI systems must be security-tested, especially if used in decision-making, public-facing platforms, or sensitive environments.

What boards should consider

- [ ] Are AI systems in our organisation developed and/or deployed with **security embedded from the start**?
- [ ] Have we carried out **penetration testing** on any systems using AI and if so, what did we learn? What corrective actions have been taken?
- [ ] Are the implications of future re-training understood and managed securely?
- [ ] Do developers and engineers working with AI **formally commit to safe and secure development practices**?
- [ ] Do we have a mechanism for **logging and responding to AI security incidents** including false positives, misuse, or breach attempts?
- [ ] Do we have **monitoring and reporting processes** in place if suppliers identify security problems that affect our systems?

Boards must ensure that AI systems are designed not only to deliver value, but to be resilient, trustworthy, safe and secure throughout their lifecycle.

Test and evaluate systems and remove from use if unintended impacts or harms are discovered

Before any AI system is deployed it must be rigorously tested to ensure alignment with the organisation's ethics, safety and/or responsibility frameworks, performance expectations, and legal obligations.

Testing outcomes should be evaluated and recommendations provided. Based on this, the board may:

- Approve full implementation
- Adjust the scope of use
- Delay or veto deployment if critical concerns are identified

This principle also requires that the board maintain ongoing oversight of deployed AI systems, to ensure they continue to operate safely, fairly, and consistently over time. If a system is found to introduce or reinforce bias, cause harm, or deviate from its original purpose, there must be a clear and tested mechanism to pause, remediate, or retire it.

This is not only a matter of operational safety - it is a signal of governance integrity. Responsible use of AI includes a commitment to course correction when things go wrong.

The same expectation applies to externally sourced AI systems, such as tools integrated into HR platforms, CRM systems, or supply chain software. **Procurement processes must include due diligence on ethical standards, explainability, and bias controls.**

What boards should consider

- [ ] Does our governance framework require that all AI systems are **tested for ethical, legal, and technical compliance** before deployment?
- [ ] Are we satisfied that the **independent review committee (or equivalent) plays a meaningful role** in reviewing evaluation outcomes and influencing deployment decisions?
- [ ] Does the board retain **final accountability for implementation**, including the ability to **reject or reverse deployment** where appropriate?
- [ ] Do we have a **clearly defined process** for removing or pausing AI system use if bias, harm, or unintended outcomes are discovered - even post-launch?
- [ ] When procuring AI-enabled tools from third parties, are **ethical and safety requirements embedded in our procurement** and contract management processes?

An organisation's credibility in using AI responsibly is not only built on how it adopts and/or designs systems, but also on how it responds when things go wrong. The board's role is to ensure there is a culture of vigilance, supported by clear pathways for action.

Review systems, policies and governance practices regularly

The governance of AI doesn't end at deployment. AI systems must be subject to regular review cycles to ensure they continue to serve their intended purpose, operate safely and fairly, and remain in alignment with the organisation's ethical commitments and risk appetite. Likewise, AI governance frameworks require regular re-assessment against KPIs, and broader regulatory and technological developments.

These reviews should assess:

- The system's ongoing alignment with its original scope and purpose
- Whether the system is producing unexpected impacts, outcomes, or drift
- Whether data inputs, models, or use contexts have changed over time
- Whether affected stakeholders are still being treated fairly and transparently

Where deviations are found, corrective actions should be:

- Timely - appropriate to the severity of the issue
- Well-documented - with traceable decisions
- Subject to governance

Reviews should also include a Data Use Impact Assessment (DUIA): an emerging concept drawing on a framework similar to Data Protection Impact Assessments (DPIAs) which assesses data beyond technical performance.

Boards should champion the use of assurance tools such as:

- Algorithmic Impact Assessments
- Ethical audit frameworks
- Certification mechanisms (e.g. BSI Kitemark, ISO standards)
- Ongoing human-in-the-loop oversight where needed

What boards should consider

- [ ] Are we conducting **regular, proportionate reviews of all AI systems** in use?
- [ ] Do these reviews include a Data Use Impact Assessment, not just technical metrics?
- [ ] Are systems still operating in line with the declared purpose?
- [ ] Do we have thresholds and escalation pathways for when systems drift, degrade, or become unreliable?
- [ ] Is there a human-in-the-loop process for systems involved in high-stakes or automated decision-making?
- [ ] Are we making use of **independent assessment or third-party assurance tools** where appropriate?

By building regular, structured review into AI governance, boards demonstrate not only regulatory readiness but responsibility, strategic foresight, and long-term stewardship.

Appendix A

Acronyms, definitions and other terms used in this guide

Some of the following terms are also covered in more depth and detail in the IoD Glossary of Science, Innovation and Technology terms.

AI (Artificial Intelligence)
AI can be defined in many ways. Broadly, it is an umbrella term to describe a range of technologies and approaches when computers or machines are built to do tasks that usually require human thinking, like learning or decision-making.

AI governance
AI governance in the context of this document is the system by which an organisation manages its (development and) use of AI systems. This includes a wide range of governance structures, policies, practices and other mechanisms that guide AI use and management. The broader topic explores the ways in which organisations can evolve existing corporate governance approaches and introduce new strategies to realise the benefits of AI systems while addressing the risks.

Algorithmic Impact Assessment
A review that checks what effects an AI or algorithm might have on people, fairness, and safety.

Black Box AI
AI systems whose internal workings are not transparent and are hard to understand; you can see the outputs but not really how they came about.

CIO (Chief Information Officer)
An executive responsible for overseeing an organisation's information technology strategy and ensuring that IT systems effectively support business goals. In the UK they are sometimes known as the Director of IT.

CISO (Chief Information Security Officer)
A senior executive responsible for protecting an organisation's information and technology systems from security threats. The CISO sets the overall cybersecurity strategy, ensures compliance with relevant laws and standards, and leads incident response efforts to safeguard sensitive data and maintain trust.

CSR (Corporate Social Responsibility)
How a business acts responsibly and gives back to society, not just making money.

D&I / DEI / EDI (Diversity, Equity, and Inclusion)
A set of principles and practices aimed at creating a fair and equitable environment where everyone has equal opportunities, regardless of their background or personal characteristics.

DPIA (Data Protection Impact Assessment)
A check done to see how a new project might affect people's privacy and how to reduce any risks.

DSIT (Department for Science, Innovation and Technology)
A UK government ministerial department that aims to accelerate innovation, investment and productivity through science; ensure that new and existing technologies are safely developed and deployed across the UK; and drive forward a digital government.

ESG (Environmental, Social and Governance)
A framework for assessing a company's impact and performance in the following three areas: Environmental, Social, and Governance.

EU AI Act
A comprehensive set of legally binding rules in force in Europe to help make sure AI is safe, fair, and well-governed.

FCA (Financial Conduct Authority)
The financial regulatory body in the UK that enables a fair and thriving financial services market for the good of consumers and the economy.

GDPR (General Data Protection Regulation)
A set of legally binding rules from the EU that protect people's personal information and how it's used by companies.

GenAI (Generative AI)
A type of AI through which (often multimodal) models can generate new content, including audio, code, images, text, simulations and videos, based on the data on which they were trained.

Human-in-the-loop
When a person is still part of the decision-making, even where AI is being used.

ICO (Information Commissioner's Office)
The UK's data protection watchdog, which helps make sure organisations use people's information properly.

ISO (International Organization for Standardization) A group that creates shared rules and best practices for doing things safely and consistently worldwide.

ISO/IEC 42001
A new official standard that helps organisations manage AI responsibly across its entire lifecycle.

ISO/IEC 5259
A set of standards to help organisations make sure their data is clean, traceable, and trustworthy.

KPI (Key Performance Indicator)
A way to measure how well a company or team is doing something important.

LLM (Large Language Model)
A type of foundation model trained on vast amounts of text to understand and generate human-like language and other modes of content. LLMs are built on machine learning: specifically, a type of neural network called a transformer model.

ML (Machine Learning)
A subfield of AI which learns from data and improves over time without being explicitly programmed to do so.

Model drift
The degradation of machine learning model performance due to changes in data or in the relationships between input and output variables. In other words, when an AI system's accuracy or fairness starts to change over time.

Ofcom (Office of Communications)
Ofcom is the UK's communications regulator, overseeing television, radio, telecoms, and postal services to ensure they operate in the public interest.

PDF (Portable Document Format)
A type of computer file that looks like a printed page and is easy for people to read, but not always easy for computers to work with automatically.

RAG (Retrieval Augmented Generation)
A method in AI where a model doesn't rely only on what it was trained on, but also pulls in up-to-date or external information from a database or document store during use. This helps it give more accurate and relevant answers, especially when dealing with specialised or time-sensitive topics.

RFID (Radio-Frequency Identification)
A technology that uses tiny chips and radio waves to store and send information, like the way modern passports can be scanned at border controls.

SSDLC (Secure Software Development Lifecycle)
A method for building tech systems with safety and privacy baked in from the start.

SLM (Small Language Model)
A smaller version of the larger (and better-known) large language model (LLM). SLMs have fewer parameters and require much smaller training datasets, optimising for efficiency and better suiting them for deployment in environments with limited computational resources.

Synthetic Data
Artificial data created algorithmically, designed to mimic real-world data, and retaining the underlying statistical properties of the original data on which it is based.

Tripwire (in AI governance)
A tool akin to a smoke detector for AI, that watches for warning signs that an AI system is capable of generating substantially harmful outputs such as designs and plans for weapons, industrial espionage and so on.

Appendix B
Additional resources

Australian Institute of Company Directors and Human Technology Institute (2024) A Director's Guide to AI Governance.

Kavanagh, J. Doing AI governance (Substack)

Lord Clement-Jones (2022) Chapter 54 'Artificial Intelligence'. From: 'Effective Directors: The right questions to ask'.

Lundblad, N. B. and Kokko, P. AI for Boards (Substack)

Patel, O. Enterprise AI Governance (Substack)

Thomas, M. (Blog)

Walker, L. Enterprise AI Executive (Blog)

Please scan the QR code to share your feedback on the paper

About the Expert Advisory Group on Science, Innovation and Technology

The Science, Innovation and Technology Expert Advisory Group was established by the IoD's Policy team to help tap into the expertise of IoD members on the key issues for UK directors, providing insight from those who have substantial front-line experience.

About the Policy Team at the Institute of Directors

The IoD's Policy Team provides the Institute and its membership with insight and thought leadership on all aspects of business policy. Through its regular interactions with government and politicians, the IoD Policy Team is an influential voice for business leaders in the UK and beyond.