AI Governance and Risk Management:

A Practical Guide for Leaders

"Leaders are expected to make decisions on AI without a practical, non-technical governance framework; this booklet fills that gap."

November 2025 Edition

Louise Wilmarsdotter Bjork

This booklet is guidance, not legal advice. It reflects the state of AI and regulation in late 2025 and must be adapted to local laws, sector requirements and your organisation's own risk assessments. The governance model is intended for global use. Examples and references draw mainly on EU, US and international standards, and should be adapted to local laws and sector-specific regulation.

At a glance

Artificial intelligence is now part of everyday organisational life. It is in productivity tools, search, customer service, analytics and security. Leaders do not need to become data scientists, but they do need a clear view of what AI is, what it is not, and how to govern it sensibly.
This booklet sets out practical guidance for boards and senior managers who want to use AI confidently without closing their eyes to the risks.

What AI actually is

Modern AI systems analyse data, recognise patterns, generate content and make predictions. Different modalities matter:

- Machine learning and deep learning models predict and classify.
- Generative AI and large language models (LLMs) create new content in text, images, audio, video and code.
- Multimodal models work across several input types at once.
- Agentic systems use models to execute tasks and trigger actions.
- Rule based systems follow explicit logic.
- Most real applications are hybrids of several of these.

The type of system in use drives the type of governance required. An LLM summarising documents does not carry the same risk profile as an agent that can fire off emails, change records or release payments.

What AI is not

AI is not conscious, not self directed and not a decision maker. It has no intention, no ethics and no sense of your organisation beyond the data and instructions it receives at a given moment.
Its outputs are shaped by three things:
The data it was trained on.
The way it has been configured and integrated.
The prompt or context you provide.
If any of those are weak, the output will be weak, no matter how confident it sounds.
AI does not know whether what it says is true, fair or legal. It is a sophisticated pattern engine, not a mind. Responsibility for decisions remains firmly with humans.

The main risk categories

The booklet distinguishes three levels of risk:
Company specific risks
These affect your own organisation and can be managed internally: incorrect outputs, bias in results, more scalable cyberattacks and the misuse of sensitive data.
Social risks
These include amplification of stereotypes, deliberate manipulation through synthetic content and uneven labour market disruption. Your organisation can easily become either part of the solution or part of the problem.

Structural and long term risks
These relate to dependency on opaque systems, accelerated capability in sensitive scientific areas and the use of AI in military and security contexts. The focus here is on governance, transparency and maintaining human control, not on science-fiction storylines.

Governance essentials

Good AI governance is not exotic. It is an extension of existing responsibilities in data protection, risk, security and ethics.
Key elements include:
Clear leadership ownership and an AI steering group with technology, legal, security, risk, HR, finance and business lines represented.
Data classification: knowing what is highly sensitive, moderately sensitive and public.
The use of enterprise grade tools, not consumer interfaces, for any confidential or regulated data.
Retention and access policies that define how long AI related data is stored, where it sits and who can see it.
Human review for any decision with legal, financial or human impact.

Safe use in practice

The booklet provides concrete expectations for teams:
Verify important outputs, especially where factual accuracy, fairness or safety are involved.
Use multi factor authentication and zero trust principles on all accounts that connect to AI tools.
Label AI generated content internally so that colleagues know when extra scrutiny is required.
Train staff to understand fabrication, bias and data handling rules, and make it socially acceptable to challenge AI results.

Policies, checklists and oversight

To turn principles into routine practice, the booklet offers:
An AI Acceptable Use Policy outline.
A vendor assessment checklist for procurement and due diligence.
Incident response basics for misuse, data exposure or compromised accounts.
Three practical checklists: governance needs assessment, risk mitigation actions and board/procurement oversight points.
These are meant to be short, operational tools, not theoretical essays.

Preparing for what comes next

AI capability will continue to develop. Your governance model must be designed to adapt rather than to be finished. That means regular review of policies, training and technical configurations, along with steady investment in internal competence.

The central message is simple: treat AI with respect, keep humans firmly in charge, and do not outsource your judgement. With that mindset, AI becomes a powerful ally rather than an uncontrolled experiment.

Table of Contents

Introduction

Artificial intelligence is already woven into everyday organisational life. It sits inside office tools, search, customer service platforms, analytics and security systems. Most leaders did not ask for it, but they are now accountable for the way it is used.

This booklet is for boards, executives and senior managers who want to treat AI neither as magic nor as background noise, but as something that deserves clear rules and adult supervision. It does not try to turn you into a data scientist. It gives you enough structure to ask the right questions, set sensible boundaries and recognise when something is drifting into risk.

The focus is practical governance: how modern AI systems behave, where the main risks lie, how to handle data, what "human oversight" should mean in real workflows, and what policies, structures and metrics you ought to have in place. The examples draw mainly on European regulation, US and international frameworks, but the principles are global.

Use this booklet as a working tool. Read the executive summary, scan the "Key points for leaders" boxes and then decide which sections matter most for your organisation today. The technology will move. Regulation will evolve. Your governance will need to be updated. The aim here is not perfection, it is a clear, robust starting point you can build on.

1. Understanding Modern AI Systems

1.1 What AI Is

Artificial intelligence refers to technical systems that perform tasks normally associated with human cognition, such as recognising patterns, interpreting information, generating content or making predictions. Modern AI does not understand the world. It identifies statistical relationships in large amounts of data and produces outputs that fit those patterns.

Different categories of AI support different types of tasks. Leaders benefit from distinguishing them, because the risks, strengths and governance needs are not the same.

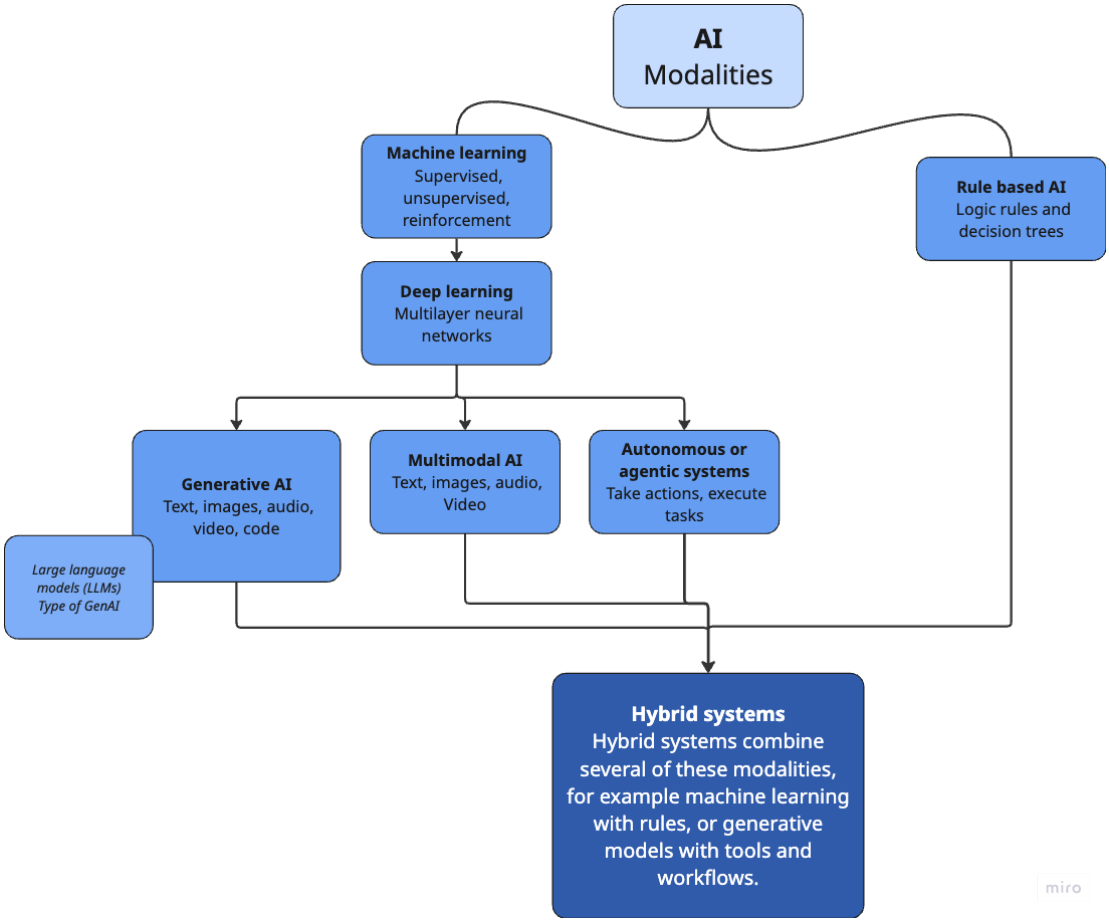

Figure 1 . This model distinguishes classic rule-based AI from learning systems, shows deep learning as a subset of machine learning, and places generative, multimodal and agentic systems as application layers built on top. Large language models are treated as a type of generative AI. Hybrid systems combine several of these modalities in real-world applications.

A. Machine Learning (ML)

Machine learning systems learn from data to make predictions or classifications.
Key types include:

- ***Supervised learning***: learns from labelled examples to predict outcomes.
- ***Unsupervised learning***: finds patterns or clusters in unlabelled data.
- ***Reinforcement learning***: learns through trial and error in a controlled environment.

Common uses: forecasting, fraud detection, demand planning, recommendation systems.

Risk note: performance depends entirely on data quality and training objectives.

B. Deep Learning

A specialised subset of machine learning that uses multi-layered neural networks.
These models excel at:

- image recognition,

- speech recognition,
- complex pattern detection.

They require large datasets and significant computing power.

C. Generative AI (GenAI)

Generative AI produces new content based on patterns learned in training data.
It can create:

- text,
- images,
- audio,
- video,
- code,
- simulations and design concepts.

Examples include LLMs, image generators and multimodal models.

This is the most visible form of AI today because it works through natural language and feels accessible to non-technical users.

Risk note: highly capable but prone to fabrication if ungrounded.

D. Large Language Models (LLMs)

LLMs are a category within generative AI.
They process and generate natural language by predicting the next most likely word or sequence.

Capabilities include:

- drafting text,
- summarising documents,
- translating languages,
- answering questions,
- analysing information,
- supporting decision preparation.

LLMs do not understand meaning in the human sense. They detect linguistic patterns.

E. Multimodal AI

Systems that work across several modalities at once, such as:

- text,
- images,

- audio,
- video,
- code,
- sensor data.

These models can interpret a diagram, answer with text, produce a visual, and analyse audio all within the same system.

This is where many enterprise tools are heading.

F. Autonomous or Agentic Systems

These systems combine AI models with the ability to take actions or execute tasks.
They can:

- search information,
- run sequences of prompts,
- trigger workflows,
- interact with software tools,
- perform operations without continual human input.

Agents increase efficiency but also increase operational risk because they initiate actions rather than simply generating content.

G. Rule-Based AI

The oldest form of AI, based on logic rules and decision trees.
Still widely used in:

- compliance checks,
- workflow automation,
- eligibility systems,
- simple chatbots.

Risk is low but inflexible.

H. Hybrid Systems

Most real systems combine several forms of AI.

Example:
A customer service AI might use:

- an LLM for language,
- ML for classification,
- rules for compliance,
- an agent to trigger an action,

- retrieval for factual grounding.

Understanding these components helps leaders set appropriate governance approaches.

Why This Matters for Leaders

Different modalities imply different governance needs:

- LLMs: need verification and grounding.
- Agents: need oversight, limits and monitoring.
- ML systems: need quality data, fairness controls and documented training.
- Multimodal models: need careful handling of sensitive inputs, including images and voice.
- Rule-based systems: need logic reviews and scenario testing.

Strong governance begins with an accurate understanding of what kind of AI you are dealing with.

1.2 What AI Is Not

AI is not conscious, not self directed and not a decision maker. It has no intention and no judgement. It produces outputs based on statistical likelihood, which is why human oversight is essential.
AI does not know whether what it says is true, fair, legal or aligned with your policies. It has no sense of context beyond the data and instructions it is given in that moment. It cannot see the wider organisation, the politics in the room or the consequences of a bad decision.
The quality of its output depends on three things: the data it was trained on, the way it has been configured, and the prompt you provide. If any of those are weak, the result will be weak, no matter how confident the wording sounds.
Treat AI neither as a threat nor as a saviour, but as a powerful assistant that can speed up thinking and execution, while still needing a human brain, a policy framework and a sense check.

Key points for leaders:

No agency: AI cannot take responsibility. People do.
No built in truth check: it does not automatically verify facts unless you design a system that forces it to.
No persistent understanding of your organisation: unless you provide context or integrate with your systems, it has no idea how your company works.
Probabilistic, not exact: it will sometimes be wrong, even when it sounds certain.
Easily steered: whoever writes the prompt has real influence over the outcome, for better or worse.

- You do not need to be a modeller, but you must know which type of AI you are approving: generative, LLM, agentic, rule based, hybrid and so on.
- Different modalities imply different risks and different controls. An LLM that drafts text is not the same as an agent that can execute actions in your systems.

- AI is a pattern machine, not a mind. It has no intent, no ethics and no concept of your organisational reality unless you provide context.
- Never delegate responsibility for a decision to “the AI”. Responsibility always remains with a human.

2. Core Categories of AI Risk

2.1 Company Specific Risks

These risks directly affect organisations and can be managed internally.

- Incorrect outputs: Models still produce fabrications, though usually subtle. Verification remains essential.
- Bias: Outputs can reflect hidden patterns in training data or design choices.
- Scalable cyberattacks: AI enables faster, more targeted phishing and identity misuse.
- Data misuse: The risk comes from misconfiguration, use of consumer tools, or careless handling of sensitive data.

Key points for leaders

- The most expensive AI failures are usually not technical. They are governance failures: unverified outputs, unmanaged bias, or poor processes.
- Treat fabrication as a design constraint, not a surprise. Critical workflows must include verification and, where possible, technical grounding.
- Bias is not an edge case. Assume it is present and require teams to monitor and mitigate it.
- Cyber risk increases with every new integration. Insist on security reviews for any AI system that connects to mail, identity, finance or core platforms.
- Data misuse is far more likely to come from people using the wrong tools than from an attack on the model provider.

2.2 Social Risks

These require awareness beyond the company boundary.

- Amplified prejudices: Stereotypes can be scaled rapidly by automated content.
- Manipulation: Synthetic identities and tailored narratives can influence public opinion.
- Labour disruption: Tasks shift quickly, creating stress and inequality unless managed responsibly.

Key points for leaders

- Your organisation is now a potential amplifier of social narratives, for better or worse. Content policies and review processes are part of AI governance.
- Assume that synthetic content and automated influence campaigns exist in your information environment. Train your teams to recognise them.
- Plan for workforce transition. AI will alter roles and tasks. Silence from leadership will be filled by speculation and fear.

- Social responsibility is not a “nice to have”. Reputational damage from careless AI use travels faster than any press release.

2.3 Structural and Long Term Risks

These relate to systemic dependency rather than dramatic science fiction.

- ***Accelerated scientific capability***: AI lowers the expertise barrier in sensitive fields.
- ***Military use***: Reduced human oversight can increase escalation risks.
- ***Dependency:*** Organisations may rely on opaque systems they cannot fully explain or control.

Key points for leaders

• Focus first on real structural risks: dependence on opaque systems, concentration of power and loss of human oversight, before worrying about science-fiction scenarios.
• Map where AI already sits in critical processes such as finance, customer service, infrastructure, supply chain and security.
• Require that any system involved in high-impact decisions has a clear escalation path to a human who can overrule it.
• Engage with sector regulators and industry bodies early rather than waiting for fully finalised rule books.

3. Modern Data Governance Principles

3.1 Data Classification

Leaders should categorise data clearly:

- Highly sensitive data such as personal records, legal material, R&D and source code.
- Moderately sensitive data.
- Public or low sensitivity content.

3.2 Use Enterprise Tools

Sensitive data should only be used in enterprise grade AI environments that, at a minimum, provide:

- No training on user inputs.
- Encryption at rest and in transit.
- Access controls and audit logs.

3.3 Retention and Access Policies

Organisations should define how long AI related data is stored, who may access it and why.

Key points for leaders

- Data classification is the foundation. If people do not know what is sensitive, they cannot handle it correctly.
- Mandate enterprise grade tools for any processing of confidential or regulated data. Consumer tools are for sandboxing, not for client work.
- Ask explicitly how long data is retained, where it is stored and who can see it. If no one can answer quickly, governance is not yet in place.
- Align AI data rules with existing privacy, security and compliance frameworks so that teams are not drowning in conflicting policies.

4. Safe Use Practices for Teams

4.1 Human Review

Any decision with legal, financial or human impact requires human verification. AI can support judgement but cannot replace it.

4.2 Zero Trust Security

Use minimal privilege access, multi factor authentication and continuous monitoring.

4.3 Internal Labelling

AI generated content should be labelled so colleagues understand where extra care is required.

4.4 Training and Awareness

Teams should understand fabrication, bias and data handling rules. A little scepticism goes a long way.

Key points for leaders

- “Human in the loop” is only meaningful if the human is trained, has time to review and feels authorised to say no.
- Make it clear that checking and challenging AI outputs is expected, not a sign of mistrust or incompetence.
- Reward good judgement, not just speed. If people are punished for slowing down to verify, they will quietly stop doing it.

- Invest in basic AI literacy for all knowledge workers. It is cheaper than cleaning up after a poorly understood tool causes trouble.

5. Governance Structures for AI Adoption

5.1 Leadership Responsibilities

Leaders must:

- Approve AI principles.
- Allocate resources to safe adoption.
- Ensure regulatory alignment.

5.2 AI Steering Committee

A cross-functional group should oversee:

- Risk assessments.
- Vendor reviews.
- Usage policies.
- Incident response.

5.3 Documentation and Transparency

Organisations should document:

- Purpose of each system.
- Data sources.
- Limitations.
- Known risks.

Key points for leaders

- Governance must be visible. If people cannot name who owns AI policy and who sits on the steering group, it might as well not exist.
- Cross-functional is not optional. You need technology, legal, risk, security, HR, finance and business lines at the same table.
- Documentation is a defence. When something goes wrong, the question will be “what did you know and what did you put in place”. Keep that answer ready.
- Start small but formal. Even a lightweight set of principles and a simple register of AI use cases is better than an untracked sprawl of experiments.

6. Policy Templates for Organisations

6.1 Acceptable Use Policy

Defines appropriate tasks, prohibited activities and escalation processes.

6.2 Vendor Assessment Checklist

Includes scrutiny of security guarantees, retention rules, compliance and safeguards.

6.3 Incident Response Procedures

Covers steps for handling misuse, errors or data exposure.

Key points for leaders

- Policies are living tools, not PDFs in a folder. Require that every major AI initiative maps to the Acceptable Use Policy, vendor checklist and incident process.
- Keep templates short and operational. If a manager cannot apply a policy to a specific case in under five minutes, it needs simplification.
- Align sanctions and incentives. If breaching policy has no consequence, or following it creates only friction, it will be ignored.
- Revisit templates after the first real incident. The best improvements usually appear once you have seen how things fail in practice.

7. Practical Checklists for Leaders

7.1 Governance Needs Assessment

A tool to identify strengths and gaps in:

- Strategy and purpose.
- Accountability.
- Policies and standards.
- Technical environment.
- Risk awareness.
- Compliance.
- Monitoring and improvement.

7.2 Risk Mitigation Checklist

Covers:

- Data protection.
- Human oversight.
- Security controls.
- Technical safeguards.
- Transparency.
- Incident readiness.

7.3 Board and Procurement Oversight

Focuses on:

- Strategic alignment.
- Vendor due diligence.

- Regulatory compliance.
- Ethical considerations.
- Risk assessment.
- Transparency.
- Security and misuse prevention.
- Human oversight.
- Implementation readiness.
- Lifecycle management.

8. Preparing for the Future

AI adoption is continuous. Leaders should review governance quarterly, monitor technical developments and adjust training. Policies should adapt as regulations evolve. A responsible organisation remains curious, careful and willing to challenge its own assumptions.

Key points for leaders

- AI capability will not stand still. Your governance model must be designed to adapt, not to be "finished".
- Build internal capacity, not just vendor relationships. You need people who can ask the right questions, not only contracts with impressive logos.
- Watch how your peers are organising themselves: boards, AI councils, ethics committees, risk forums. Borrow what works.
- Keep talking about AI in plain language. The more it is treated as normal, the easier it becomes to surface concerns early instead of hiding them.

9. Regulatory Anchors and External Frameworks

This booklet is not a legal guide, but it should sit comfortably beside the main frameworks that are shaping organisational practice.

9.1 EU AI Act: How This Governance Model Helps

The EU AI Act introduces a risk based regime with strict obligations for high risk systems, including requirements on risk management, data governance, documentation, logging, transparency, human oversight and robustness. Digital Strategy EU+1
For providers and deployers of high risk AI systems, obligations include: maintaining a risk management system throughout the lifecycle, ensuring data quality and relevance, keeping logs, providing clear instructions for use, enabling human oversight and achieving appropriate levels of accuracy, robustness and cybersecurity. Artificial Intelligence Act+2Artificial Intelligence Act+2
How this booklet lines up:

- Sections 2 and 3 support risk management, data governance and human oversight.
- Sections 5 and 6 support documentation, instructions for use and incident handling.
- The checklists in section 7 give a starting point for assessing whether an internal use case might drift into “high risk” territory and therefore deserves legal review.

Key points for leaders

- You do not need to memorise every article. You do need to know which of your systems might be “high risk” and who in your organisation owns the compliance roadmap.
- Treat this booklet as the governance layer. Use it to frame questions for your legal, compliance and risk teams, not as a substitute for them.
- If you operate in or with the EU, ask explicitly which AI systems will need to meet EU AI Act requirements and how your current controls map to Articles 9–15 and 26.Artificial Intelligence Act+1

9.2 NIST AI Risk Management Framework: A Common Language

The NIST AI Risk Management Framework (AI RMF) is becoming the default risk language in many organisations. It describes four core functions across the AI lifecycle: Govern, Map, Measure and Manage.NIST+2I.S. Partners+2

- ***Govern***: establish risk management culture, roles, policies and processes.
- ***Map***: understand the context, intended use, stakeholders and potential impacts of a system.
- ***Measure***: assess risks, performance and impacts with suitable metrics.
- ***Manage***: prioritise and treat risks, monitor systems and adjust controls over time.

How this booklet lines up:

- Sections 5 and 6 largely sit in ***Govern***.
- Sections 2, 3 and 7 help with ***Map*** and ***Measure***.
- Sections 4 and 8 relate to ***Manage***, particularly human oversight and continuous adjustment.

Key points for leaders

- You do not have to “implement NIST” in full to benefit from the vocabulary. Using Govern, Map, Measure, Manage as headings in your own materials is often enough to align teams.
- If your risk or security teams already work with NIST frameworks, ask them to place your AI work inside that structure instead of reinventing parallel processes.
- The NIST Generative AI profile adds extra detail for systems that produce content. It is worth knowing that it exists, even if you delegate the implementation.NIST+2AI Governance Library+2

9.3 ISO/IEC 42001: AI Management Systems

ISO/IEC 42001 is the emerging international standard for AI management systems. It describes how to establish, implement, maintain and improve an organisation wide system for responsible development and use of AI.lasso.security+3ISO+3UNIDO+3

Typical requirements include:

- an AI policy and clear objectives;
- governance and risk management processes;
- impact assessment and lifecycle management;
- competence and training;
- supplier and third party oversight;
- continuous improvement.

How this booklet lines up:

- Your governance structures, policies and checklists form part of an AI management system in the ISO sense.
- If your organisation already works with ISO standards (9001, 27001 and so on), ISO/IEC 42001 can be integrated into that familiar pattern, with this booklet as a practical companion.

Key points for leaders

- Decide whether ISO certification is strategically useful. In some sectors it will be expected. In others it is a "nice to have".
- Even without certification, ISO/IEC 42001 is a useful benchmark when asking whether your governance is complete enough.
- Treat standards as guardrails, not as an excuse to switch your brain off. You still own the real world consequences.

10. Worker and Stakeholder Participation

Governance is not only a board exercise. The people who use, are evaluated by, or are affected by AI systems see risks that policies on paper will miss.

10.1 Involving Workers

Under the EU AI Act, deployers of high risk systems must inform workers that such systems are in use, ensure input data is appropriate, assign human oversight and monitor operation, including acting when risks emerge.Artificial Intelligence Act+1
Even outside formal "high risk" categories, basic good practice includes:

- explaining to staff where AI systems are used in recruitment, performance assessment, scheduling or safety;
- giving people a way to challenge or appeal AI assisted decisions;
- involving staff representatives or works councils in impact assessments for significant systems.

10.2 External Stakeholders

For customer facing or citizen facing systems, it is worth:

- testing early prototypes with real users, not only with technical teams;
- providing clear explanations in plain language where AI plays a material role in outcomes;
- capturing feedback and complaints specifically about AI mediated processes, not mixing them silently into generic customer service queues.

Key points for leaders

- Trust is earned locally. If people feel AI is being "done to them", resistance and quiet workarounds will follow.
- Worker and stakeholder input is not only a fairness issue. It is a risk control. They are often the first to spot harmful behaviours, poor data or unintended consequences.
- Make sure accountability lines are clear: who is responsible for listening, for changing systems when feedback signals harm and for reporting difficult findings upwards.

11. Metrics and Signals That Governance Is Working

If you cannot see whether governance is functioning, you are managing by hope. You do not need a vast dashboard, but you do need a handful of indicators.

11.1 Basic Coverage Metrics

- Percentage of relevant staff who have completed AI literacy and data handling training.
- Number of AI use cases recorded in an internal register.
- Proportion of those use cases that have at least a basic risk and impact assessment attached.

11.2 Risk and Oversight Metrics

- Number of incidents and near misses involving AI systems, by severity and business area.
- Frequency of human overrides or corrections in critical workflows.
- Number of policy exceptions granted, and why.
- Time between detection of a serious issue and implementation of a mitigation.

These are in line with the kind of outcomes NIST associates with "Measure" and "Manage" in the AI RMF and with the monitoring expectations in the EU AI Act for high risk systems.AI Act+3NIST AI Resource Center+3Medium+3

11.3 Value and Adoption Metrics

Governance that only measures risk will quietly fall to the bottom of the priority list. Track value as well:

- Documented time savings or quality improvements in pilot projects.
- Number of teams that have moved from experimentation to stable, governed use.
- Staff sentiment: whether people feel AI tools help them do better work or simply add pressure.

Key points for leaders

- Choose a small set of metrics you can actually review regularly, rather than an encyclopaedia that no one reads.
- Use trends, not single numbers. Spikes in incidents, sudden drops in overrides or a sharp rise in “shadow tools” are all useful signals.
- Treat metrics as prompts for questions, not as a score to be gamed.

12. Sector Specific Adaptation

This booklet is intentionally sector-neutral. The principles apply in most environments, but the emphasis changes.
Examples:

- ***Health and life sciences***: stronger focus on clinical risk, informed consent, traceability and regulatory approvals.
- ***Financial services***: close alignment with existing model risk management, anti money laundering rules and conduct supervision.
- ***Public sector and humanitarian work***: particular care for fundamental rights, inclusion, and the impacts of errors on vulnerable populations.
- ***Industrial and safety critical sectors***: integration with existing safety management systems and incident reporting frameworks.

Key points for leaders

- Start with this generic model, then layer on sector rules, local law and organisational culture.
- Ask each business unit to translate the governance principles into their own language, use cases and regulatory environment, and to bring back the gaps.
- When in doubt, design for the most demanding context you operate in. It is usually easier to relax controls than to improvise them under regulatory pressure.

Conclusion

AI can increase efficiency and support innovation, but only when it is paired with robust governance, sound data practices and informed human oversight. Leadership sets the conditions for safe and effective use. This booklet offers a practical baseline for organisations, now linked to the main external reference points such as the EU AI Act, the NIST AI Risk Management Framework and ISO/IEC 42001, and supported by checklists, metrics and questions you can use in boardrooms, steering groups and procurement.

It does not remove the need for legal advice, technical expertise or worker participation. Instead, it gives you a clear structure to organise that work: understanding what kind of AI you are dealing with, deciding how far you want to go, and knowing what “good enough for now” looks like in your context and sector.

In the end it still comes down to disciplined common sense. Treat AI with respect. Design clear boundaries. Verify its outputs. Listen to the people affected by it. And never outsource your judgement.

Sources and further reading

This booklet is designed to be practical, not exhaustive. For leaders and teams who want to go deeper, the following references provide the main external anchors for AI governance and risk management as of late 2025.

Regulation and standards

- ***EU Artificial Intelligence Act, Regulation (EU) 2024/1689***
  Final text published in the Official Journal on 12 July 2024. Provides the risk based regulatory framework for AI in the EU, including requirements for high risk systems on risk management, data governance, documentation, logging, transparency, human oversight and robustness. EUR-Lex+2EUR-Lex+2
- ***EU AI Act Explorer and consolidated text***
  Independent but widely used resources that present the structure and main obligations of the Act in accessible form, including overviews of risk categories and obligations for deployers. Artificial Intelligence Act+2Artificial Intelligence Act+2
- ***ISO/IEC 42001:2023, Artificial intelligence – Management system***
  The first international standard for AI management systems. Sets out requirements for AI policy, risk management, lifecycle management, impact assessment, supplier oversight and continual improvement, and is increasingly used as a baseline for AI governance programmes. ISO+1

Risk management frameworks

- ***NIST AI Risk Management Framework (AI RMF 1.0)***
  Developed by the US National Institute of Standards and Technology as a voluntary framework to "Govern, Map, Measure and Manage" AI risks. Widely adopted as a reference for organisational AI risk management and trustworthiness. Securiti+3NIST+3NIST Publications+3
- ***NIST AI 600-1, Generative AI Profile (2024)***
  A profile of the AI RMF focused on generative AI. Identifies risks that are novel to or amplified by generative systems and proposes concrete actions across governance, data protection, security and content integrity. clearyiptechinsights.com+3NIST Publications+3NIST AI Resource Center+3

Safety institutes and frontier AI work

- ***UK AI Security Institute (formerly AI Safety Institute)***
  Government body tasked with independent testing and evaluation of advanced AI models, and with developing methods such as safety cases and frontier model

evaluations. Its publications illustrate how external evaluation and benchmarks for high capability models are evolving. TIME+4GOV.UK+4GOV.UK+4

- ***Frontier AI Safety Commitments (AI Seoul Summit, 2024)***
  Political commitments by major AI companies and governments on safety, evaluations and incident reporting for frontier models. Useful context for boards that want to understand how global debates on safety and security are moving. GOV.UK+2Frontier Economics+2

These references are not required reading for every manager, but they are useful anchor points for legal, risk, security and compliance teams, and they explain why the governance model in this booklet looks the way it does.