ARTICLE

IM_

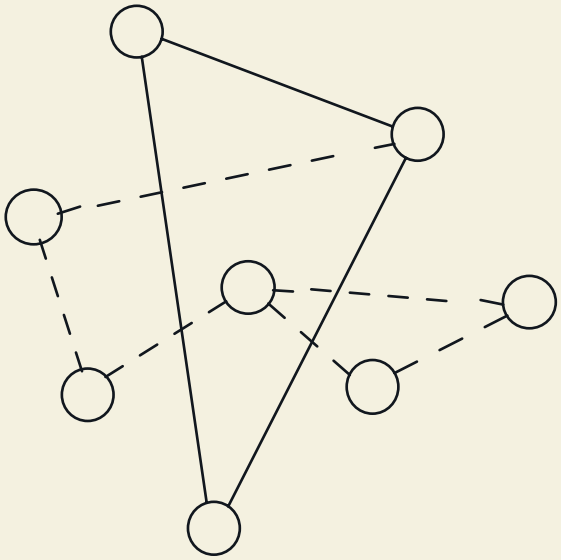
Governing agentic AI

Agentic AI does not ask permission. It reasons, decides, and acts on our behalf, at machine speed.

So, the question shifts from whether the model is accurate to whether we can keep control, accountability, and trust.

From intelligence to autonomy: where agentic AI fits

Traditional artificial intelligence (AI) refers to machine learning and statistical models that operate on structured data to automate, predict, and optimise decisions. These systems identify patterns and improve outcomes in areas such as forecasting, risk management, and operational efficiency.



Generative AI (GenAI) extends these capabilities using large language models that handle general inputs and outputs. It can analyse and create content from unstructured data such as text, images, or audio. An assistant (sometimes used interchangeably with agent) is a model configured to perform a specific task through prompt engineering and data access.

Agentic AI systems give models a degree of autonomy and access to a toolbox, so they can plan, call tools, and complete multi-step tasks. They range from a single tool-augmented model through to networks of agents that coordinate with each other. As autonomy rises, AI stops being a tool inside a process and becomes a participant in it. That is what makes deliberate control, and not just capability, the real design question.

The shift to agentic AI

Agentic AI is already here, making decisions, taking actions, and operating across workflows with growing autonomy. The change is not only technological. When AI moves from inside a process to acting as a participant in it, the assumptions that governed earlier systems no longer hold.

The opportunity is real, and it starts with technology. But there are no use cases without operations that can be trusted. An agent that can act is also one that can act wrongly, at speed and at scale. The question leaders face is less about model accuracy and more about control, accountability, and evidence. This is where governance earns its place. Not as a brake on ambition, but as the discipline that lets value move into production safely.

In our view, governance is a management discipline. It keeps AI development and use aligned with business strategy, a defined risk appetite, and the regulation that applies, so innovation can accelerate within boundaries everyone can see.

In the following, we outline how we think about AI governance and the model we use to make it concrete.

What our research tells us: AI adoption is racing ahead of governance

What this means for the model ahead

Investment and ambition are not the bottleneck. The organisations pulling ahead are those that pair AI with governance: clear ownership, human oversight where it matters, and evidence that controls are working.

Today, governance is too often experienced as friction: more than half say security and compliance slow them down, and only a small minority say it actively helps them move. The opportunity is to flip that, so governance becomes the thing that gives teams the confidence to scale rather than the thing that holds them back.

The question is less whether to invest and more whether the operating model can carry autonomy safely.

Adoption is an ongoing process, scaling is not

Ambition is high and the tools are in place, but the discipline around them is not. Only about one in five leaders has an enterprise-wide AI roadmap, while a further fifth has none at all and a third describe theirs as still in development. The intent is real; the operating model lags.

Spending intent confirms the commitment: nearly three-quarters of respondents expect to spend more on AI in the year ahead, and almost none expect to spend less. The appetite is not in question.

And the gap shows. For more than half of respondents, security and compliance slow AI down or block it outright, rather than enabling it. Most have already killed or deprioritised initiatives in the past two years, and the reasons are rarely the technology: unproven ROI, weak leadership focus, and data or compliance constraints lead the list.

Governance now separates those who scale from those who stall

41% say their AI policies feel disconnected from day-to-day work, and a further fifth have no meaningful policies yet[1]

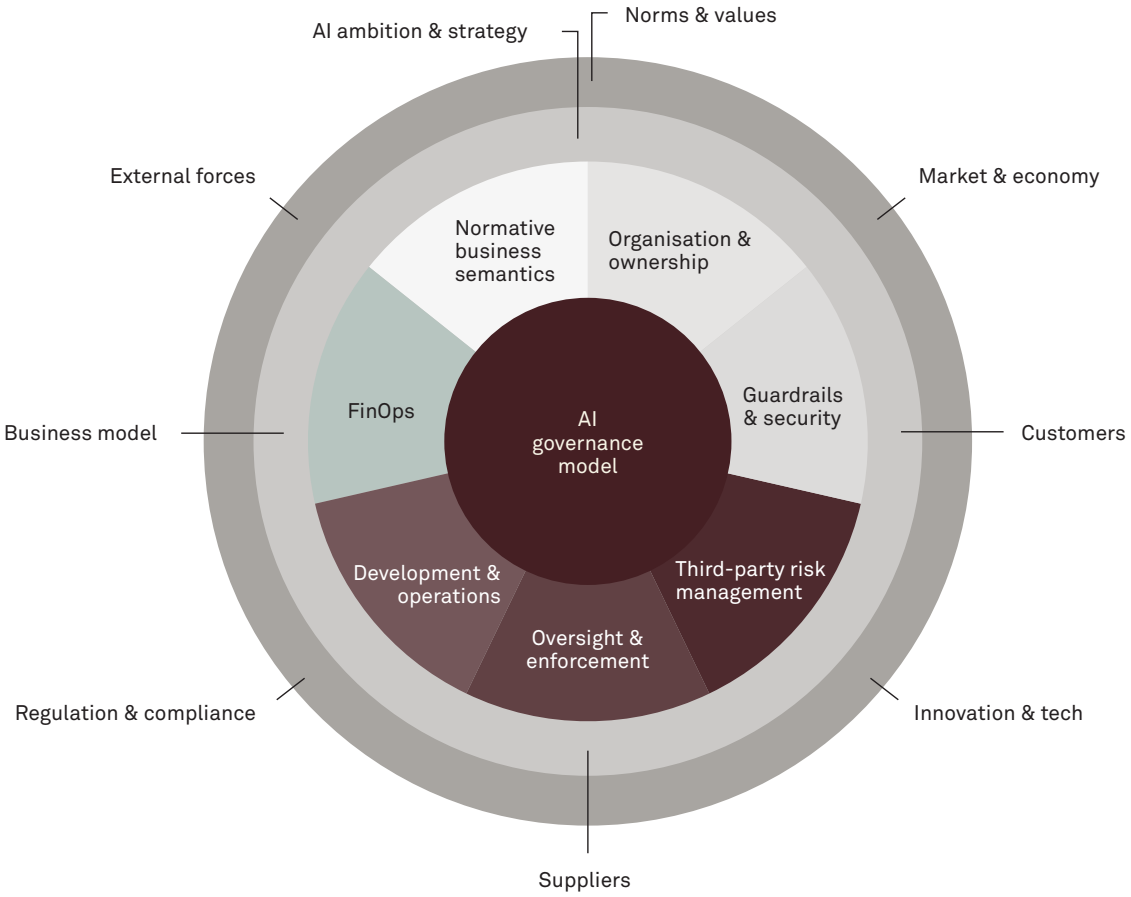
28% say the key decisions on AI investment, risk, and priorities are made with no real governance forum, more than for any formal body[1]

Across our survey, the constraint is rarely the model. It is the ownership and oversight around it.

Source: The State of AI Survey (2026), preliminary results

AI governance works as one model, not a stack of disconnected controls

AI governance is a management discipline that keeps AI development and use aligned with strategy, a defined risk appetite, and the regulation that applies. The model brings seven dimensions into one structure, framed by the external forces that shape any organisation, so AI can be scaled within clearly defined risk boundaries.



The seven dimensions that turn governance from policy into practice

The seven dimensions are interdependent. Below we outline why each one matters, and what an organisation gains when it is designed well.

The foundation

Normative business semantics translates regulation and ethical commitments into enforceable standards: risk appetite, acceptable use, model and documentation requirements. These are living assets, versioned, owned, and reviewed like the systems they govern, not a static document. It is the foundation every other dimension stands on, the agreed definition of what good looks like.

Organisation and ownership turns those rules into a living system and organisation structure. It names an owner for every AI system, sets up the governance bodies, and makes the three lines of defence work in practice. Rules without owners are process without authority, so this is what makes someone answerable.

Guardrails and security are where governance physically meets the system. Input and output controls, least-privilege access, and layered defences bound what each system can reach or do, while monitoring catches drift, fairness, and explainability failures before customers feel them. Without them, the other layers stay theoretical.

Oversight and enforcement is what makes the rest credible. First-line controls own monitoring at the point of use; the second, third, and fourth lines independently challenge what they see; and incident management resolves issues and feeds the learning back. Regulators want evidence that controls have fired, not just structures on paper.

Running it well

Third-party risk management matters because most AI capability runs on vendors. When a model fails, is biased, or is withdrawn, you remain accountable to regulators and customers. Outsourcing the technology does not outsource the risk, so vendor and shadow AI need inventory, contracts, and monitoring of their own.

Development and operations is the build-and-run spine. A single intake triages every use case by risk, prompts and models are version-controlled with gated releases, and the infrastructure makes each system reproducible. Without it, AI defaults to chaos; with it, AI becomes a managed system rather than a sequence of pilots.

FinOps closes the loop on economics. An agent does not answer once; it loops, reasoning, calling tools, and carrying context forward, so cost scales with how many turns a task takes, and a vaguely scoped request can cost many times a well-scoped one. That makes cost a governance question, not a budgeting one: a spending cap tells you that you overspent, not why. The control belongs in the platform, funding the right use cases, bounding consumption by design, and tying spend to a named outcome. Measured well, the unit that matters is cost per completed outcome, not cost per call.

Each dimension depends on the others. A gap in any one quietly undoes the rest, which is why we treat them as a single integrated model rather than a collection of separate controls.

The seven dimensions only create value when they reinforce one another

What turns a set of policies and tools into governance that actually holds?

There is no single control that makes AI safe. Governance works as a system. The model brings seven dimensions into one structure: the normative framework, organisation and ownership, guardrails and security, third-party risk, oversight and enforcement, development and operations, and FinOps. Each depends on the others, and a gap in one quietly undoes the rest.

That is why we treat them as one model rather than a collection of committees and tools. The aim is a model that is documented, scalable, and auditable, where innovation can accelerate within clearly defined risk boundaries.

Rules need owners

A normative framework sets the rules, principles, and standards: risk appetite, acceptable use, accountability, model standards. It is the foundation everything else stands on. But rules without owners are just process without authority. Organisation and ownership turns them into a living system: who owns each AI system, how the governance bodies work, how the three lines of defence operate, and how accountability is enforced.

Without it, policies exist but nobody owns them, and models reach production with no accountable party. Put simply, the normative framework says what good looks like, and organisation and ownership makes someone answerable for it.

The seven dimensions only create value when they are credible

Guardrails meet the system

Input and output guardrails, access rights, and security bound what each AI system can reach, share, or do. Monitoring and evaluation catch drift, fairness, or explainability failures before customers feel them. Guardrails are where governance physically meets the system. Without them, the other layers are theoretical.

A policy does not stop an over-permissioned agent from leaking data, and a committee does not catch a malicious instruction arriving in a document at three in the morning. Together they separate bounded autonomy from black-box risk.

Oversight makes it credible

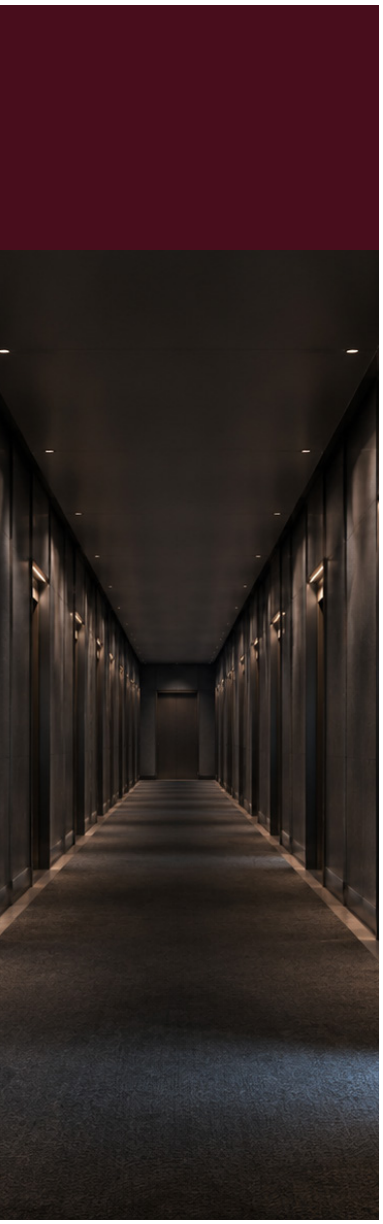
Oversight and enforcement is what makes the rest credible. Line of controls own monitoring at the point of use. In highly regulated industries such as the insurance industry, second, third, and fourth lines independently challenge what the first line sees. Incident management resolves issues and feeds the learning back into the governance operations.

Regulators do not only want to see governance structures; they want evidence that controls have fired and findings have been acted on.

Exposure and economics

Most AI capability depends on third parties. When a vendor model fails, is biased, or is withdrawn, you remain accountable. Outsourcing the technology does not outsource the risk. And because the same task can consume very different resources from one run to the next, FinOps closes the loop: funding the right use cases, catching runaway consumption before it becomes an invoice, and tying spend to a named outcome.

Measured well, the unit that matters is not cost per call, but cost per completed business outcome.



The seven dimensions only create value when they are strategic

Build and run, under control

Development and operations is the build-and-run spine. A single intake triages every use case against risk and regulation. Increasingly, control has to start at build time, not only at runtime: the context, boundaries, and constraints an agent is given before it generates or acts matter as much as the checks applied afterwards. Prompts, context, tools, and models are version-controlled, with automated tests gating each release. The data and AI infrastructure makes every system reproducible. Without this spine, AI defaults to chaos: systems built outside central IT, prompts edited live, silent vendor model swaps changing behaviour overnight.

With it, AI becomes a managed system rather than a sequence of pilots. The lifecycle runs from a controlled intake through to decommissioning, where derived data and artefacts are also retired. This is what lets autonomy scale without losing the thread of what each system is doing and why.

A thought for leadership teams

Few organisations have yet shown sustained bottom-line impact from AI. The pull towards quick wins is understandable, but short-term ROI can come at the cost of the scale and control that make the value last. Sustained impact comes from rethinking processes and governing AI deliberately – not from isolated technology investments.

Future outlook

The future of AI is impossible to predict. What is clear is that agentic AI cannot be governed with the assumptions built for generative AI.

Across industries, organisations are moving in a common direction, not because the destination is known, but because standing still is not an option. The leaders are not those with perfect roadmaps, but those who can scale with confidence because they can show how their systems are controlled.

Progress often starts with pragmatic use cases. These are not ends in themselves, but entry points: safe places to build the governance, evidence, and regulatory readiness that more autonomous use cases will demand.

The harder work lies inside the organisation. The real question is less which model to buy, and more whether you can answer, for every system acting on your behalf, who owns it, what it can reach, and how you would know if it went wrong.

Make AI explicit. Challenge whether it truly supports the strategy. Rethink your ways of working. In our experience, that is what earns an organisation the right to scale agentic AI.

Contact

For more information, please contact:

Nina M. Stegger
Implement Consulting Group
+45 22 45 08 62
nist@implement.dk

Claus Høyer Madsen
Implement Consulting Group
+45 40 70 60 35
chma@implement.dk

Kenneth Ullmann
Implement Consulting Group
+45 24 24 84 66
keue@implement.dk

Wilbur Helfer
Implement Consulting Group
+45 42 95 53 77
wihe@implement.dk

Julian Róin
Implement Consulting Group
+45 2422 3474
jurs@implement.dk