

Report

Agentic AI Governance:

The Field Guide For
Enterprise Leaders



Table of contents

1.	Introduction	3
2.	AI Inflection Part 1: AI Governance Framework and Why the Third Changes Everything	5
3.	AI Inflection Part 2: What Agentic AI Governance Requires	11
4.	AI Inflection Part 3: The Structural Components of Agentic AI Governance	16
5.	Five Tests for a True Agentic AI Governance Process	20
6.	What the Adjacent Tools Are — and Are Not	24
7.	Comprehensive AI Governance Wins Trust — and Trust Wins Everything	31
8.	The Accountability Question Nobody Is Asking Loudly Enough	35
9.	A Practical Assessment for Enterprise Leaders	38
10	About the Authors	43

Agentic AI Governance: A Field Guide for Enterprise Leaders

The AI Governance category was created in 2020 to solve a problem that the market did not yet have a name for: how do you make AI accountability contextual to the sector and use case, continuous across the full lifecycle, and comprehensive across the full scope of risk.

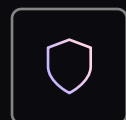
That foundational framework has navigated two major AI inflection points: the predictive AI era and the generative AI era. It is now facing a third, the most demanding yet, the Agentic AI era. This field guide explains what that demands, where adjacent tools fall short, and where the market is headed.

AI Governance: Why it matters



Accelerates Innovation

AI Governance gives enterprises the confidence to move faster — not slower. When risk is understood, documented, and managed, teams deploy AI without institutional hesitation. Governed AI is deployable AI.



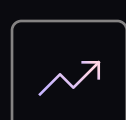
Manages Risk

AI Governance is the systematic capability to identify, assess, and manage AI risk across the full lifecycle — before deployment, at runtime, and through every change in the regulatory environment.



Ensures AI is trustworthy

Trust in AI systems is not declared — it is demonstrated through evidence. AI Governance produces the accountability record that makes trust legible to boards, regulators, and customers.



Maintains Accountability at Agentic Speed

Trust in AI systems is established when AI Governance maintains accountability that keeps up with the growing demands of agentic workflows.

The Data: The Rise of Agentic AI and the Widening AI Governance Gap

The numbers are unambiguous. Agentic AI is scaling at a rate that has no historical precedent in enterprise technology. The AI Governance

infrastructure to match it does not exist. That gap is already producing failures, abandoned projects, and accountability crises at scale.

The rise of Agentic Ai

8x

Increase in enterprise apps embedding AI agents — under 5% to 40% in one year.

Gartner, 2025

327%

Growth in multi-agent workflow usage in just four months — agents are getting exponentially more complex.

Databricks State of AI Agents, 2026

57%

Of large enterprises have AI agents running in production today.

LangChain State of AI Agents, 2026

The AI Governance gap

60% / 4%

Of organizations are deploying AI across multiple departments — but only 4% are governing it at scale.

Credo AI, State of AI Governance, 2026

80%

Of senior AI Governance leaders rate automated safeguards that block policy, violating agents as critical.

Credo AI, State of AI Governance, 2026

40%+

Of agentic AI projects at risk of cancellation by 2027 without proper AI Governance.

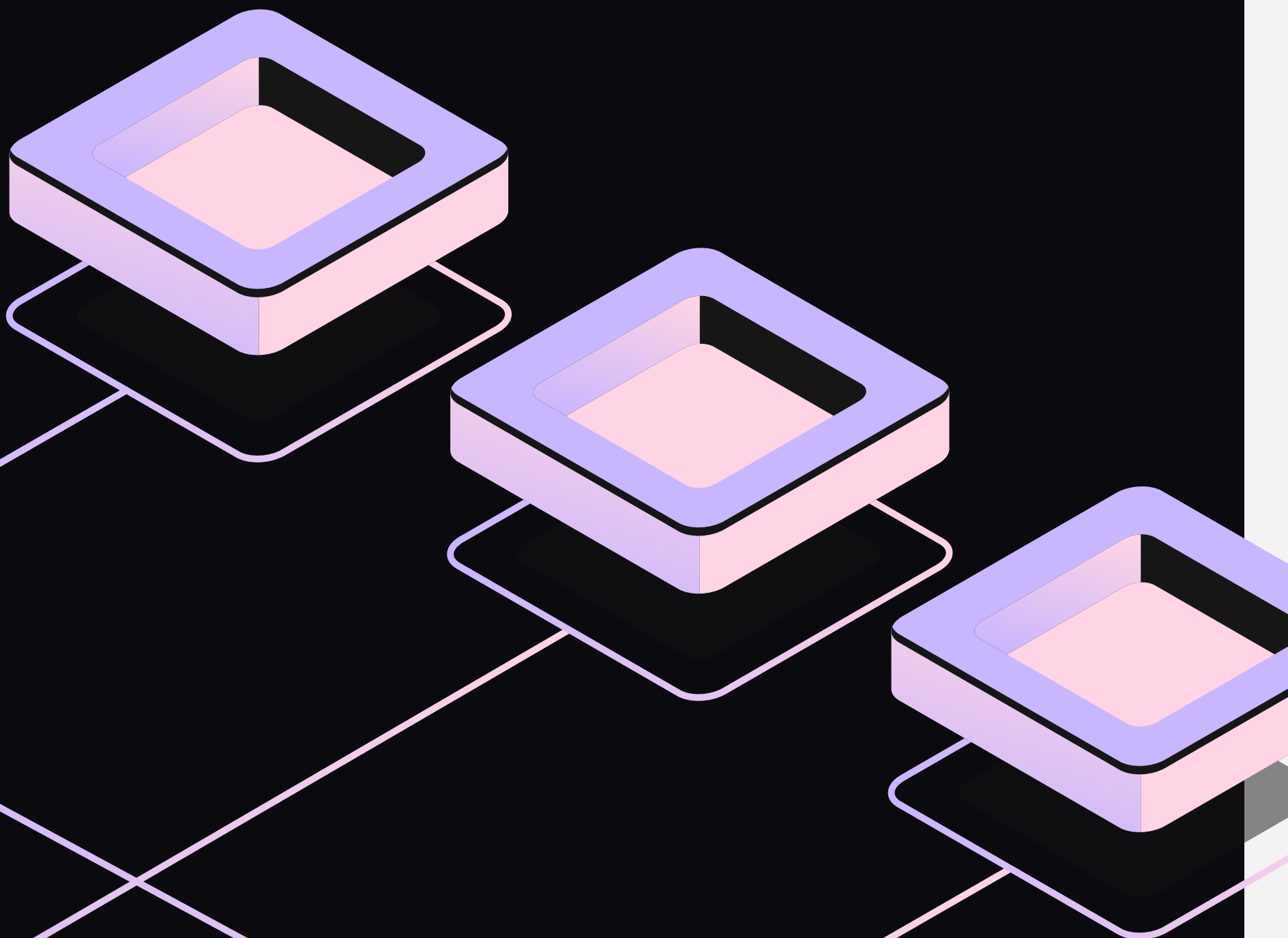
Gartner, 2025

Organizations with fully integrated AI Governance are nearly 4× more likely to report revenue growth than those still piloting — 58% versus 15%. The difference is not the model. It is not the foundation. It is accountability.

The accountability–performance link · Grant Thornton, 2026

02 • AI Inflection Part 1

AI Governance Framework and Why the Third Changes Everything



When Credo AI created the AI Governance category in 2019, the core insight was precise and deliberate.

Enterprises deploying AI needed AI Governance which is:

Contextual

Calibrated to the specific sector, use case, and risk level of each AI deployment. A credit-scoring model in financial services carries different obligations than a recommendation engine in retail. AI Governance had to reflect that.

Continuous

Not point-in-time audits, but accountability that runs across the full lifecycle: from the moment an AI system is being considered, through development, deployment, production, and decommission. The most consequential AI Governance decisions happen before a system is deployed, not after.

Comprehensive

Covering the full scope of AI risk: bias and fairness, explainability, model risk, operational risk, regulatory compliance, third-party risk, and the organizational accountability structures that connect AI systems to the humans responsible for them.

That framework was right when we built it. It remains right today. What has changed — dramatically, twice, and now a third time — is the environment it must govern.

The predictive AI era demanded AI Governance focused on model risk: was the model trained on appropriate data, tested for bias, documented, monitored for drift, and connected to the regulatory requirements of its sector. The AI Governance questions were hard but bounded — they centered on a model producing outputs that humans then acted on.

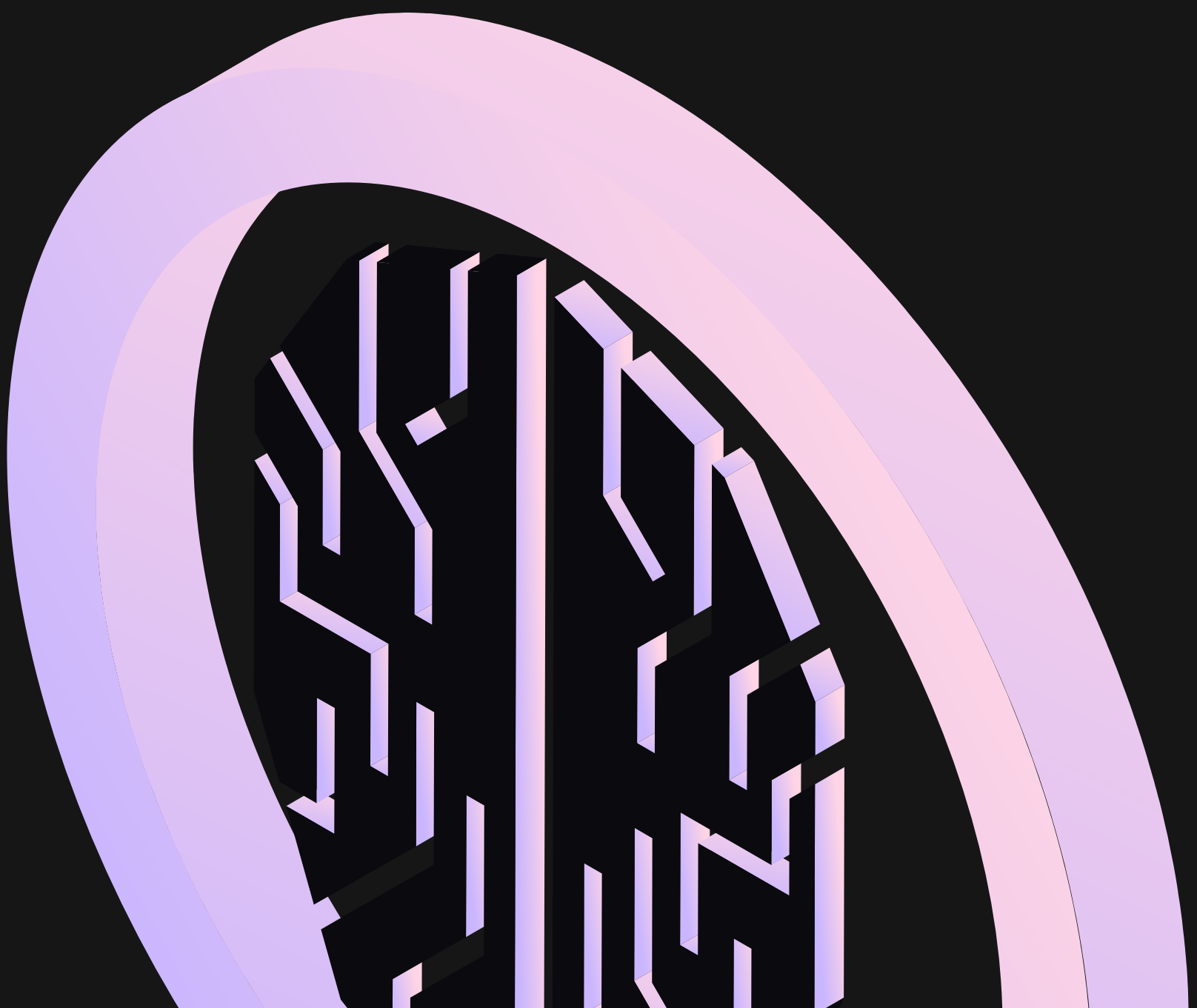
The generative AI inflection expanded the problem significantly. Foundation models introduced new risk dimensions: hallucination, content safety, the explosion of use cases built on shared infrastructure. AI Governance had to extend to cover prompt engineering, output evaluation, third-party model AI Governance, and the new compliance landscape that regulators began building in response. The contextual-continuous-comprehensive AI Governance framework scaled into this environment — but required new tooling and new policy connections to do so.

The agentic AI era is categorically different from both. AI agents do not produce outputs for humans to act on. They take actions themselves. They plan across multiple steps, call tools, access and write data, trigger downstream systems, spawn sub-agents, and operate in continuous loops — autonomously, at millisecond speed, at scale, often without a human in the loop at any point in the sequence. The system is no longer passive. It is an actor.

This changes the agentic AI Governance question fundamentally. With predictive and generative AI, the central question was: did this system produce a fair, accurate, and explainable output? With agentic AI, the question becomes: is this autonomous system operating within its sanctioned purpose, under the constraints its business context requires, in a way the organization can be held accountable for — in real time, at the speed it operates?

An agent ensures no anomalies detected, no threats flagged, no drift observed — can still be operating entirely outside the boundaries its specific business context demands. A procurement agent authorized to request quotes is not authorized to approve purchases. The difference between what an agent can do and what it should do in a given business context is not a security question. It is an AI Governance question. AI Governance helps an organization be more precise about that specification an agent's business context requires.

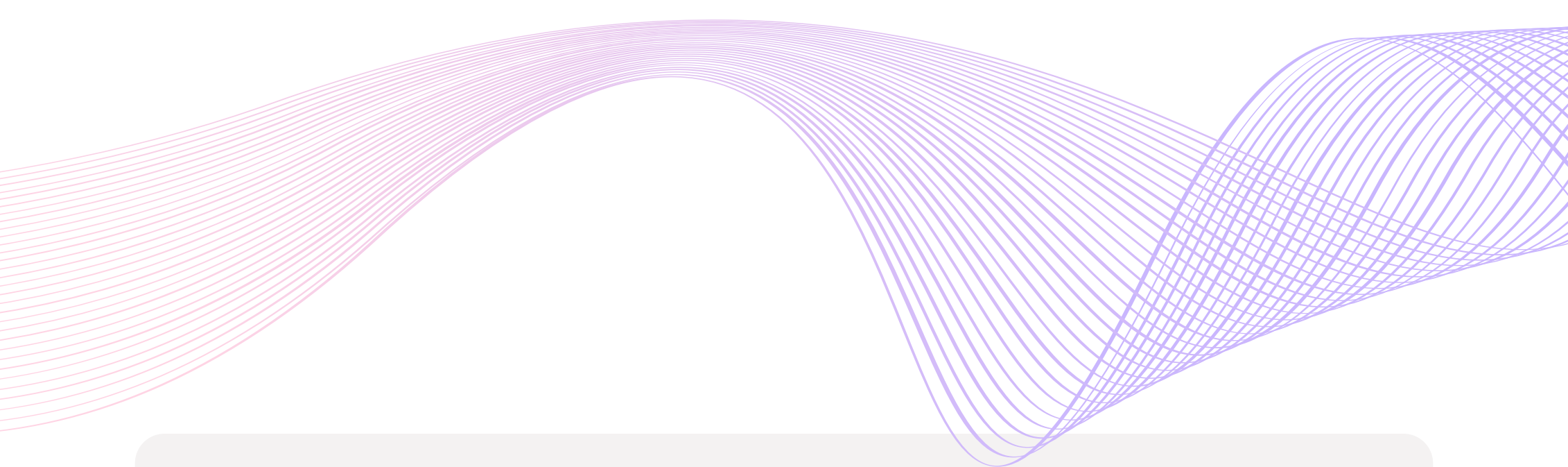
Think of agentic AI Governance as the brain, the people and process, and the enforcement tools as muscles. All of that together is Governance. Muscles without a brain do not know why they are acting, whether a given action is appropriate, or what the organization committed to when it deployed the system. Most tools currently sold as agentic AI Governance solutions are excellent muscles being asked to do the work of a brain.



The Most Important Shift in AI Governance: From Outside the System to Inside It

You cannot govern a system that takes autonomous actions in milliseconds by reviewing its outputs in a monthly audit. You cannot govern an agent that reasons across multiple steps by watching what it produces at the end of the chain. You cannot govern a system

that operates continuously by applying an AI Governance process that runs periodically. The speed, autonomy, and complexity of agentic AI systems make external AI Governance not just inadequate — they make it irrelevant by the time it activates.



The defining shift of the agentic AI era: AI Governance moves from outside the system to inside it. Not as a monitor. Not as a reviewer. As an architectural component — present in how the agent is defined, constrained, authorized, and continuously validated.

What AI Governance inside the system looks like in practice:



The Agent card

The agent card is not a document in a compliance folder. It is a live AI Governance artifact that the agent is defined against — its sanctioned purpose, its configuration, its risk classification, its applicable policy — present before the agent runs and continuously updated as it operates.



The agent constitution

The agent constitution is not a policy statement on a wiki page. It is a constraint layer embedded in how the agent reasons — the inviolable principles it cannot optimize around, present in its architecture, not just in its documentation.



The pre-deployment gate

The pre-deployment gate is not a human checklist. It is a programmatic condition — risk assessed, constitution defined, policy aligned, sanctioning recorded — that must be satisfied before the system is permitted to run. The gate is part of the deployment architecture, not an approval email in someone's inbox.



The runtime AI Governance layer

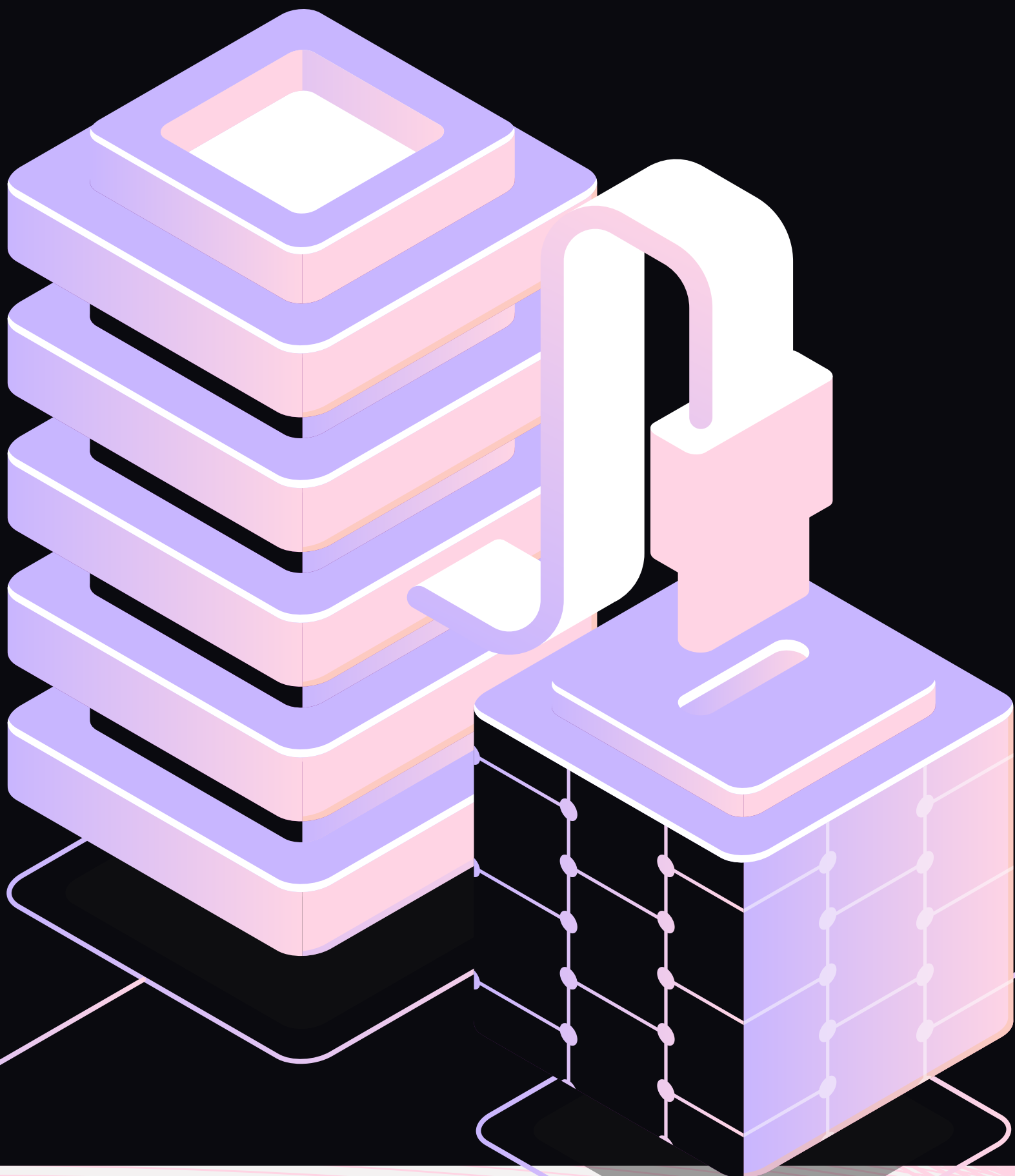
This layer is not a dashboard a human checks at the end of the week. It is a continuous validation loop running alongside the agent's execution loop — comparing actual behavior against the agent card, detecting deviations from the constitution, routing AI Governance events to the right human at the right time with the right context.

This shift

From AI Governance as bureaucratic overlay to AI Governance as architectural component — is what makes the agentic AI Governance brain categorically different from the monitoring and AI security tools adjacent to it. Those tools sit outside the system and observe it. The AI Governance brain is designed to be inside the Governance loop, present in how agents are defined, constrained, authorized, and continuously held accountable. That is not a feature distinction. It is an architectural distinction. And it is the reason AI Governance cannot be built by bolting a compliance module onto a security platform.

03 • AI Inflection Part 2

What Agentic AI Governance Requires



Agentic AI Governance is the organizational capability to ensure that autonomous AI systems operate within their sanctioned purpose.

In addition, it ensures that the autonomous AI systems are in compliance with applicable policy, configured correctly for their business context, with full accountability across their lifecycle — at the speed those systems operate.

Sanctioned purpose

The starting point. Before an AI agent operates in an enterprise environment, someone must have made a deliberate decision: this agent has this purpose and scope, is approved to do these things, with access to these systems, under these constraints. Without sanctioning, agentic AI Governance has no baseline. An agent that has never been sanctioned cannot be governed — it can only be monitored.

Business context configuration

The expression of sanctioned purpose in operational terms. Agentic AI Governance requires specifying — precisely, per interaction — what the agent is permitted to do in its specific context. What tools can it call? What data can it access? What decisions can it make autonomously versus escalate? This is an AI Governance artifact, not a technical setting — the translation of business policy into operational constraints.

Effectively, this configuration defines the identity under which the agent acts: the credentials it uses, the systems those credentials reach, and the conditions under which its permissions can be revoked in real time. Without a governed identity of its own, the agent acts under whatever account it inherited — and the audit trail cannot tell which actions were taken by the agent versus by the human or service it borrowed access from.

The agent constitution

Where configuration specifies what an agent is permitted to do operationally, the constitution defines the inviolable principles governing how it reasons and acts — regardless of task or context. Think of the difference between a job description and a code of conduct.

An agent's constitution codifies: the ethical constraints it cannot reason around, the categories of action it must never take regardless of instruction, the escalation triggers that override autonomous judgment, and the values hierarchy that governs when conflicting instructions arise. The constitution is the agentic AI Governance layer that prevents capable reasoning from becoming a liability.

This concept is emerging from AI safety research and will become a standard agentic AI Governance requirement as agentic deployments mature and regulators develop more specific expectations for autonomous system behavior.



Applicable policy

Connected to actual regulatory requirements and organizational commitments — not anomaly thresholds or threat rules. The EU AI Act requires specific documentation, risk classification, and human oversight provisions for high-risk systems. AI Governance that is not connected to these frameworks is operational monitoring with AI Governance language applied.

Full accountability across the lifecycle

Agentic AI Governance cannot activate only at runtime. The most consequential decisions happen before an agent is deployed: risk classification, model selection, data access, permitted actions, accountability ownership. Agentic AI Governance begins at the moment an agent is being considered, not the moment it is launched.

At the speed and scale agentic AI systems operate

Agentic AI Governance must itself be automated: policy codified, connected to enforcement, updated continuously. Human oversight remains critical — but it is scarce, fallible, and in many of the contexts most consequential for AI trustworthiness, simply not fit-for-purpose. The point is not only that human review cannot scale (though it cannot); it is that even where it could, AI systems will often do the work better. "Human-in-the-loop" Governance should therefore be allocated judiciously, reserved for the decisions where human judgment holds a genuine comparative advantage over AI oversight.

What the moment demands is a new abstraction layer: AI Governance systems that partner with human overseers to interpret, triage, and resolve Governance issues as they arise. The end state is not human oversight or automated oversight, but a combined human-AI system whose output is trustworthy AI deployment.



Enterprise maturity assessment

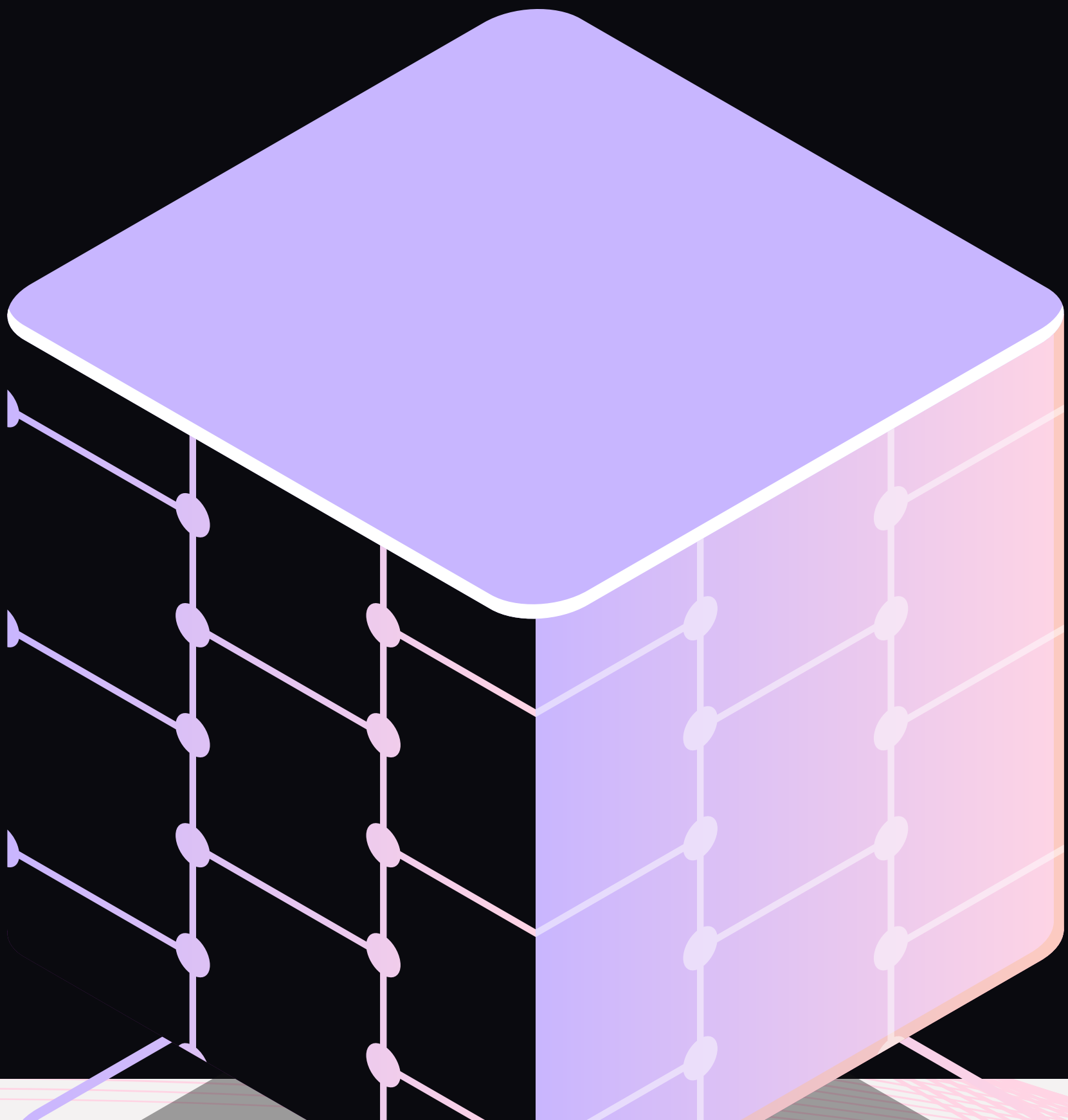
Where does your organization actually stand against the six requirements today? The table below reflects what we are seeing across the enterprise market. Most programs, assessed honestly, will find significant gaps — not because of a failure of intent, but because the agentic environment has moved faster than the AI Governance infrastructure built for the era before it.

Where Most Enterprises Stand Today

Requirement	Where most enterprises are today	Enterprise Readiness
Sanctioned purpose	Informal or absent for most agents, especially bought ones	Not Yet Production-Grade
Business context configuration	Partially addressed through technical access controls; AI Governance layer missing	Not Yet Production-Grade
Agent constitution	Almost universally absent; not yet on most enterprise roadmaps	Not Yet Production-Grade
Policy connection	Manual, periodic, not connected to deployed agents in real time	High
Lifecycle accountability	Runtime monitoring exists; pre-deployment agentic AI Governance largely absent	Moderate-High
AI Governance at agent speed	Almost universally absent; human review cycles dominate	Not Yet Production-Grade

04 • AI Inflection Part 3

The Structural Components of Agentic AI Governance



The 6 requirements translate into 6 concrete structural components.

Each is distinct. Together they constitute a complete agentic AI Governance system. The absence of any one is a meaningful gap.

1. The AI Agent Registry

The foundation of everything else — the authoritative, continuously updated record of every AI agent operating in or connected to the enterprise environment. Built agents, bought agents, agents embedded in SaaS platforms, shadow agents. Most enterprises today cannot produce this list. You cannot govern what you cannot see. The registry draws the line between what is a legitimate, sanctioned business tool and what is a shadow AI experiment, giving Governance professionals the authoritative versions to trust.

2. The Agent Card

Agent Cards function like a nutrition label for an agent, documenting not just what it is, but what it can do, what it can touch, and how autonomously it can act. An agent card is the AI Governance record for a specific agent – providing what is called “Governance-grade memory.” Documents: what the agent is designed to do, its business context configuration, the model it uses, the tools it can invoke, the data it can access, the decisions it can make autonomously versus escalate, its risk classification, the policies it must comply with, who sanctioned it, and the evidence trail of compliant operation. The Agent Card must be versioned so that for any action the agent took, you can reconstruct exactly which policies, permissions, and constraints were in force at that moment. Without that, the card is documentation. With it, the card is the authoritative record that makes accountability possible. This is the document a compliance officer references during an audit and the record a regulator requests when investigating an incident.

3. The Agent Constitution

If the Agent Governance Specification is the job description, the constitution is the code of conduct — and it operates at a different level. The constitution defines the principles an agent cannot reason around, regardless of what its task instructions say. It specifies categories of action that are never permitted, how the agent must behave when it encounters conflicting instructions, and when autonomous judgment must yield to human review unconditionally. In Credo AI's State of AI Governance report, 80% of senior Governance leaders rated automated safeguards that block policy-violating agents as a critical capability, the constitution is the layer they're asking for.

Importantly, a constitution is only meaningful if it can be enforced. That requires expressing it in checkable form, with structured rules, executable policy, and exact reference values, and matching the verification method to the claim being checked. A second language model reviewing the first is not verification; it is shared probabilistic risk. Governance-grade verification requires the right mechanism for each type of check, not a single judge applied uniformly.

4. The Pre-Deployment Gate

The formal process that must be completed before an agent is permitted to operate. Includes: risk assessment calibrated to business context; regulatory classification; business context configuration review; constitution review confirming inviolable constraints are defined; policy alignment verification; and formal sanctioning — a named approval recorded in the agent card. Without it, no one has formally accepted responsibility for what the agent does.

5. Runtime Agentic AI Governance

Continuous validation that an agent's actual behavior is consistent with its agent card and constitution. AI security asks: is this agent being attacked? Observability asks: is it performing technically? Runtime agentic AI Governance asks: is this agent doing what it was sanctioned to do, within its constitution, in the business context it was configured for? An agent can pass all AI security and observability checks and still be operating outside its agentic AI Governance boundaries.

6. The Tool Gateway

Where the pre-deployment gate determines whether an agent is permitted to operate at all, the tool gateway determines whether each action it attempts is permitted to execute. Every tool call is mediated against the agent's sanctioned permissions, business context, and constitutional constraints before reaching the system that would carry it out. Without a gateway, the same probabilistic system deciding what to do is the only thing deciding whether it is allowed to. Simply put, the pre-deployment gate is the hiring decision. The tool gateway is the approval workflow every consequential action runs through afterward. It is one of the many “muscles”, which is an important part of the Agent Governance Specification.

6 structural components in short

1. The AI Agent Registry

The authoritative record of every AI agent in your environment. You cannot govern what you cannot see.

2. The Agent Card

The Governance-grade record for each agent: what it is, what it can do, and the evidence trail behind every action.

3. The Agent Constitution

The inviolable principles an agent cannot reason around, regardless of instruction.

4. The Pre-Deployment Gate

The formal approval every agent must clear before it is permitted to operate.

5. Runtime Agentic AI Governance

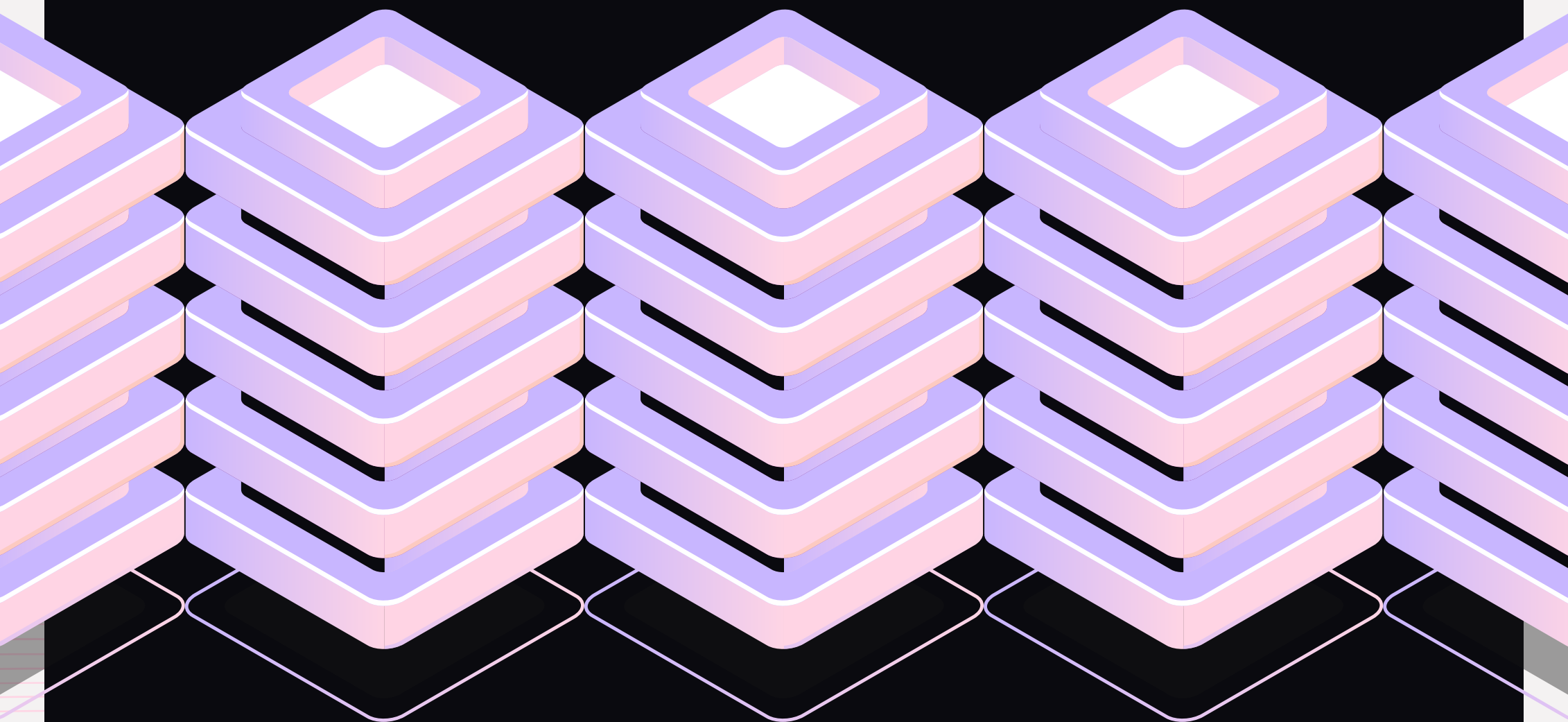
Continuous validation that an agent's actual behavior matches its sanctioned purpose.

6. The Tool Gateway

The enforcement point that checks every action against the agent's permissions before it executes.

05 •

Five Tests for a True Agentic AI Governance Process



Five Tests for a True Agentic AI Governance Process

Test 1

The Pre-Deployment Test

Does AI Governance begin before deployment? If a platform only activates once an agent is running, it is not governing the most consequential decisions. Pre-deployment risk assessment, regulatory classification, business context configuration review, constitution validation, and formal sanctioning must all exist before the agent runs.

Test 2

Accountability The Accountability Record Test

Does it hold an agent card and constitution for every agent? Ask to see the agentic AI Governance record for any specific agent. It should include sanctioned purpose, business context configuration, the agent's constitution, risk classification, evidence of least privilege, applicable policies, deployment approval, and evidence trail. If those records do not exist in structured, retrievable form, there is no agentic AI Governance — there is monitoring without a reference point.

Test 3

Policy Connection Test

Does it connect to external policies? Not custom regulatory test or anomaly thresholds, but to actual regulatory frameworks emerging for agentic AI. The frameworks below are converging on the same architectural requirements this guide describes. Can the platform demonstrate connection to them? If not, it is operational monitoring with agentic AI Governance language applied.

Framework	Description	Date
Singapore IMDA Model AI Governance Framework for Agentic AI	First jurisdiction-issued agent-specific Governance framework; risk bounding by design, meaningful human accountability, lifecycle expectations	Released January 2026
NIST CAISI AI Agent Standards Initiative + NCCoE Software and AI Agent Identity and Authorization	Federal standards work on agent security, identity, authentication, authorization, auditing, and non-repudiation	Expected Q4 2026
EU AI Act + GPAI Code of Practice (Safety and Security Chapter)	Agentic capabilities and autonomous use addressed as systemic risk dimensions; Article 50 transparency and high-risk classification enforceable from August 2026	Enforcement Aug 2, 2026
UK FCA Mills Review	Review and recommendations to FCA Board of the long-term impact of AI on retail financial services, including agentic and autonomous systems	Recommendations Summer 2026

Test 4

The Behavioral Boundary Test

Does it validate behavior against business context configuration and the constitution? This is the test no AI security or observability tool can pass. Can the platform detect when an agent is operating outside its configured business context or violating its constitutional constraints — not because it has been attacked, not because performance degraded, but because it is acting outside its agentic AI Governance boundaries?

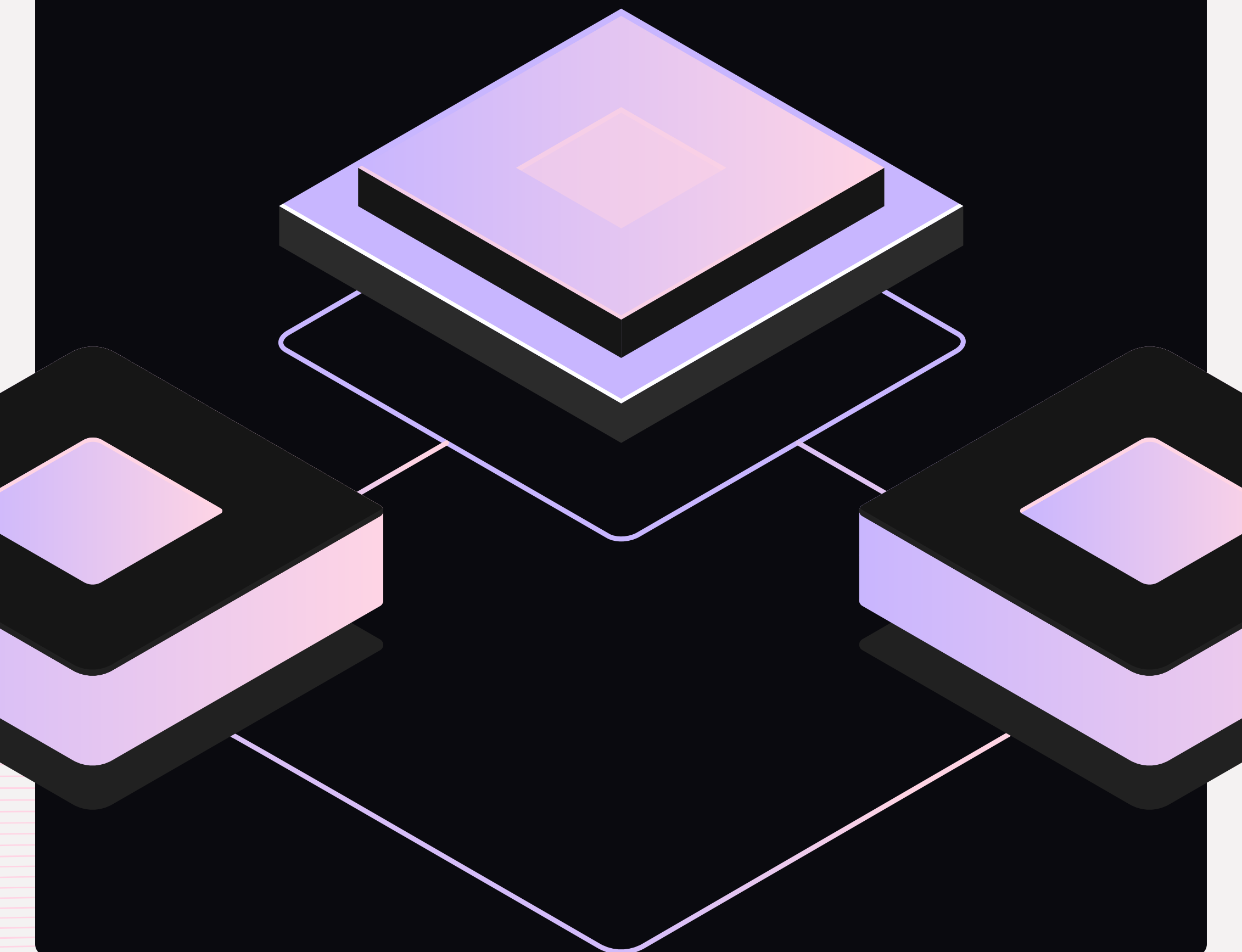
Test 5

The Bought Agent Test

Does it govern bought agents, not just built ones? Most enterprise agentic AI deployments involve vendor-built agents — embedded in Salesforce, ServiceNow, Microsoft, Workday. An agentic AI Governance platform that only works when you own the code is not adequate for the environment most enterprises are actually building into.

06 •

What the Adjacent Tools Are — and Are Not



AI Governance, AI Security, AI Observability, and Data Privacy are different jobs.

The confusion in the market comes from vendors using overlapping language. All three planes talk about 'policies,' 'guardrails,' and 'AI safety.' But they answer fundamentally different questions, report to different people, and produce different artifacts.

01/4

AI Governance

“Do we trust building, buying, or deploying this AI system?”

Value: System of record for every AI entity (use case, agent, vendor, model etc) in the enterprise across risk and compliance. Translate internal company policies and external regulations into machine readable controls.

Produces: Use case registry, risk classification, policy as code, traceable approval workflow, regulator ready evidence, and board and customer trust reports.

In short: “Measurable trust artifacts for board, customers, regulators .”

Owners: Chief AI /Data Officer, CISO (recently) , with CTO, Legal, Compliance, as stakeholders

02/4 AI Security

"Can an attacker make our agent do something bad?"

Value: Real-time defence against threats unique to AI: prompt injection, jailbreaks, tool misuse. Blocks, redacts, and quarantines at runtime.

Produces: Blocked attacks, security incidents, red-team findings

In short: "Security stops attackers. It tells you what was blocked — not whether you should have built it."

Owners: CISO · Head of AppSec, with SOC, Red Team, Cloud Security as operators

03/4 AI Observability

"Is the agent actually doing its job well?"

Value: Measurable quality of agents. Traces every reasoning step and response. Detects hallucination, drift, task failure, and regression.

Produces: Execution traces, evaluation scores, hallucination metrics, cost dashboards

In short: "Observability helps engineers fix the agent. It doesn't translate into something a regulator will accept."

Owners: VP Engineering · Head of AI Platform, with ML engineers, data science as users

04/4

Data & Privacy

"Is our data handled lawfully and responsibly?"

Value: Proof of compliance. Consent tracking, PII detection, data lineage, residency enforcement, and classification.

Produces: Consent logs, PII reports, lineage maps, residency confirmations, impact assessments

In short: "Privacy ensures lawful data handling. It doesn't tell you whether decisions are fair or auditable."

Owners: DPO • Chief Privacy Officer, with Legal, Compliance as stakeholders

The tools currently being conflated with agentic AI Governance are genuinely important.

AI Security

Answers the question: is this system being attacked, exploited, or manipulated? Runtime threat detection, prompt injection defense, adversarial input protection, unauthorized tool usage monitoring — real and necessary capabilities.

Representative vendors:

Lakera

Protect AI

HiddenLayer

Prompt Security

Zenity

Noma Security

What they do not answer:

- Was this agent sanctioned to operate?
- What business context configuration governs it?
- Does it have a constitution defining its inviolable constraints?
- What policy applies
- What is the audit trail that demonstrates accountable operation?

AI security protects the agent from external threats. Agentic AI Governance defines what the agent is permitted to do and ensures it acts within those boundaries regardless of threat.

AI Observability

Answers the question: is this system performing as expected? Model drift, output quality degradation, latency anomalies, behavior changes — operational intelligence that matters for system health.

Representative vendors:

Datadog New Relic

What they do not answer:

- Does this agent have a current agent card and constitution
- Is it classified correctly under applicable regulation
- Is it operating within its business context configuration?

A system can have excellent observability and zero AI Governance. Observability signals are inputs that agentic AI Governance needs to interpret against the agent card — they are not agentic AI Governance themselves.

Data Privacy

Answers the question: is this system handling personal data correctly? Data lineage, consent management, access controls, retention policies — essential compliance capabilities.

Representative vendors:

BigID

Securiti

OneTrust

Privacera

Privacy compliance is one dimension of AI Governance, not a substitute for it. An agent can be fully privacy-compliant and still have no AI Governance: no agent card, no constitution, no sanctioning, no accountability record.

The AI security, AI observability, and data privacy tools are the muscles of the agentic AI stack — fast, specialized, essential. The agentic AI Governance brain holds the agent card, the constitution, the policy connection, and the accountability record. It tells the muscles what to enforce and interprets what they report back. An enterprise with strong muscles but no brain has fast responses and no accountability.

07 •

Comprehensive AI Governance Wins Trust — and Trust Wins Everything



Every major technology market eventually clarifies around a single question: who does the enterprise trust with its most consequential decisions?

In cloud infrastructure, the answer was the vendor with the broadest capability and the most accountability for what ran on their platform. In cybersecurity, it was the vendor who could see the most and respond the fastest. In AI, the trust question is still being decided — and the stakes are higher than in any previous technology wave, because the decisions AI systems make affect people, not just systems.

The consolidation now underway in AI security and observability is not the story. It is the setup for the story. As AI security tools get absorbed into infrastructure platforms and observability capabilities get bundled into SaaS offerings, the muscle layer of the AI stack is becoming commoditized. Every hyperscaler, every major enterprise software vendor, every cybersecurity incumbent is building or acquiring the capability to detect threats, monitor performance, and enforce data policies at the AI layer. That is a natural market evolution.

But here is what that evolution reveals: when the muscles are everywhere, the brain is everything.

The trust problem no platform vendor can solve

Consider what an enterprise needs from its agentic AI Governance layer. It needs a system that tells the truth about every AI agent operating in its environment — including the agents built by the platform vendors the enterprise depends on. It needs a system that assesses risk honestly, not optimistically. It needs a system that connects AI behavior to regulatory requirements without a commercial interest in softening that assessment.

No platform vendor can be that system for the agents they built. Microsoft cannot be the neutral agentic AI Governance brain for Copilot agents. Salesforce cannot govern Agentforce impartially. A cybersecurity platform that acquired an observability tool is not positioned to produce the accountability artifacts that comprehensive agentic AI Governance requires — its commercial interest is in the security layer, not in the AI Governance record that sometimes finds the security layer insufficient.

This is not a criticism of those platforms. It is a structural observation about what trust requires. Trust in AI Governance is only meaningful when the governing entity has no stake in the outcome of what it governs. An independent, purpose-built agentic AI Governance platform is the only entity architecturally capable of providing that assurance. And as AI agents become more autonomous, more consequential, and more embedded in enterprise operations, that assurance becomes the most valuable thing in the stack.

Comprehensive AI Governance Is the Compounding Asset

The enterprises and the vendors that understand this early will compound their advantage in ways that are difficult to close later. For enterprises: an AI Governance program built correctly now — with a complete agent registry, agent cards, agent constitutions, pre-deployment gates, and runtime validation connected to real policy — becomes the institutional memory of trusted AI deployment. Every agent sanctioned, every risk assessed, every incident documented becomes part of an accountability record that grows more valuable with every regulatory requirement, every customer inquiry, and every board question. That record cannot be built retroactively.

For the market: the vendor that owns comprehensive agentic AI Governance owns the most trusted position in the AI stack. Not the most visible. Not the most technically impressive. The most trusted. And in a world where AI systems are taking actions that affect customers, employees, financial outcomes, and societal impact at scale, trust is not a soft value. It is the hardest competitive moat in the market.

The company that can tell an enterprise board, a financial regulator, a healthcare authority, or a government procurement officer, "here is the complete AI Governance record for every autonomous system operating in your environment, here is the evidence that each one was sanctioned appropriately, governed continuously, and aligned to the policy that governs your organization" is the company that wins the agentic AI era. Not on features. Not on integrations. On trust.

The company that wins comprehensive AI Governance **wins trust**. And in the agentic AI era, trust is the only asset that compounds.

08 •

The Accountability Question Nobody Is Asking Loudly Enough



With agentic AI, an agent makes a sequence of autonomous decisions.

This includes planning, reasoning, acting, all of which produces an outcome affecting a customer, a financial transaction, a hiring decision, a medical workflow. Who is accountable? The organization that deployed the agent.

Not the model provider. Not the platform vendor. The organization that made the decision to put that agent into operation, with that configuration, under that constitution, with those constraints.

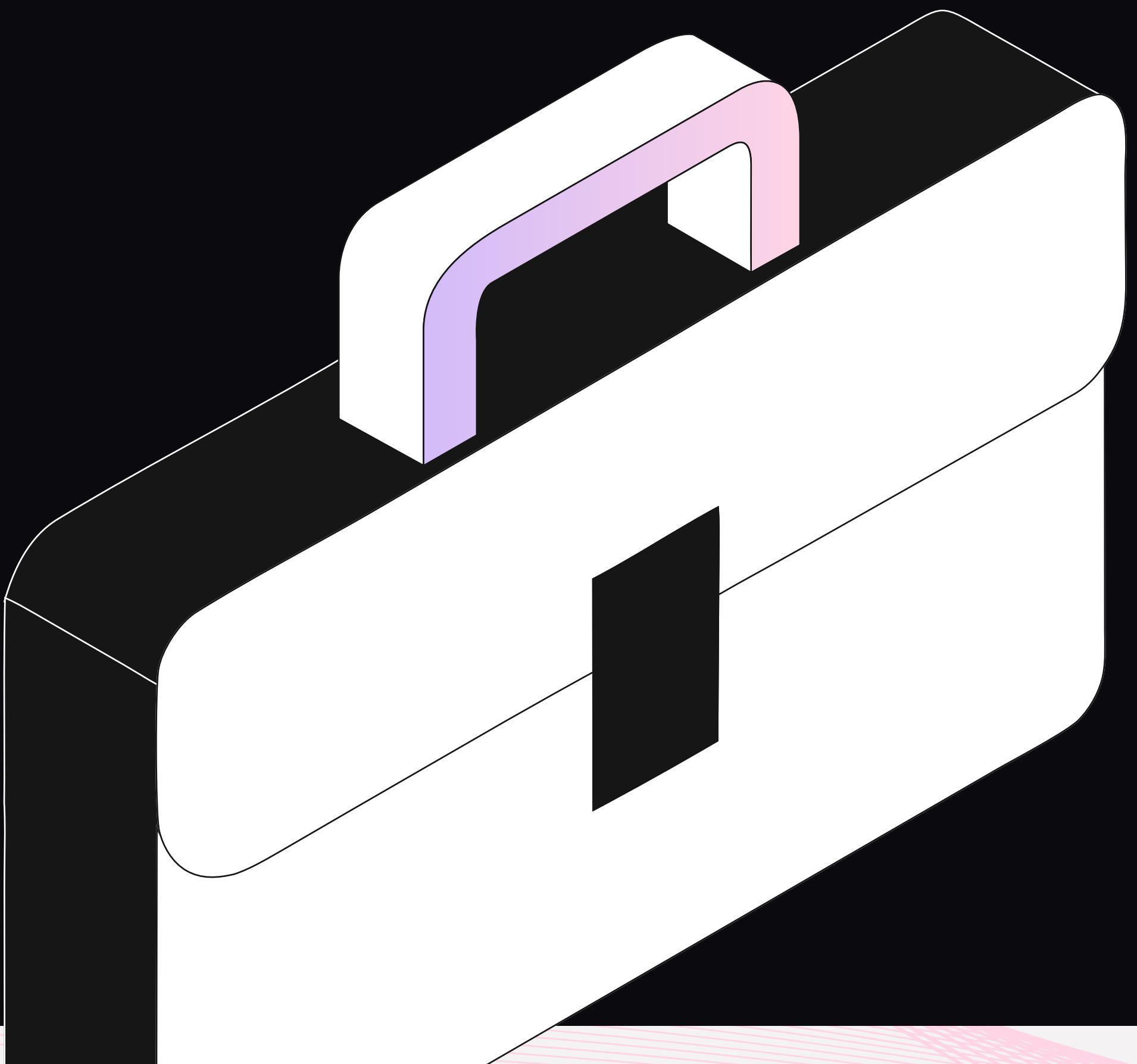
That accountability cannot be satisfied by runtime monitoring logs. It requires the full agentic AI Governance record: the registry entry, the agent card, the constitution, the pre-deployment approval, the business context configuration, the runtime validation history, that demonstrates the organization acted responsibly in deploying an autonomous system that then acted in the world.

This is the accountability gap that exists in most enterprises today. And it is the gap that regulators, in multiple jurisdictions, are now beginning to require organizations to close on enforcement timelines that are already running.

The European Commission's enforcement of the EU AI Act's general-purpose AI obligations begins August 2, 2026, with high-risk system requirements activating on the same date. The US federal banking agencies issued SR 26-2 in April 2026 and signaled a forthcoming interagency request for information specifically addressing banks' use of generative and agentic AI. Singapore's Infocomm Media Development Authority published the world's first agentic-specific Governance framework in January 2026. NIST's Center for AI Standards and Innovation closed its request for information on securing AI agent systems in March 2026, with an AI Agent Interoperability Profile expected in Q4 2026. The UK Financial Conduct Authority's Mills Review will deliver recommendations to its Board in summer 2026. The accountability gap is not a future regulatory problem. It is a present enforcement reality.

Part 09 •

A Practical Assessment for Enterprise Leaders



Most enterprise programs, assessed honestly against these questions, will find significant gaps.

That is not a failure of intent. It is a reflection of how quickly the agentic environment has outpaced the AI Governance infrastructure built for a previous era.

Inventory

Can you produce a complete, current list of every AI agent operating in your environment — built, bought, embedded, and shadow? Do you know who sanctioned each one and under what conditions?

Agent Cards

Does a complete agent card exist for every agent? Does it include business context configuration, risk classification, applicable policies, accountability owner, and an evidence trail of compliant operation?

Agent Constitutions

Does every agent in your environment have a defined constitution — the inviolable principles governing how it reasons and acts, regardless of task instructions? If not, your agentic AI Governance has no floor.

Pre-Deployment Gate

Is there a formal agentic AI Governance gate every agent must pass before operating? Does it include risk assessment, regulatory classification, configuration and constitution review, and a named sanctioning approval?

Runtime Validation

Is your runtime agentic AI Governance validating behavior against agent cards and constitutions — not just monitoring for anomalies and threats? Can it detect when an agent operates outside its sanctioned business context?

Regulatory Readiness

For each agent, can you demonstrate on demand that it meets applicable regulatory requirements, based on current evidence — not documentation produced at deployment and never updated since?

About Credo AI

From Pioneering AI Governance to Leading in the Agentic Era

Credo AI created the AI Governance category in 2019 with a specific and deliberate architecture: AI Governance that is contextual to the sector and use case, continuous across the full AI lifecycle, and comprehensive across the full scope of risk. That was not a product decision. It was a foundational insight about what accountability for AI actually requires — and it has proven correct through every major shift the market has navigated since.

Through the predictive AI era, Credo AI built the infrastructure of that conviction: the AI registry, policy-as-code, lifecycle AI Governance connected to actual regulatory frameworks, evidence-based accountability that holds up when boards and regulators ask the hard questions. Credo AI worked with some of the world's largest and most regulated organizations — in financial services, healthcare, defense, and technology — to build AI Governance programs that are operational, not aspirational.

Through the generative AI inflection, Credo AI extended the framework: into foundation model risk, output evaluation, third-party model AI Governance, and the EU AI Act compliance requirements that the generative era made urgent. The contextual-continuous-comprehensive AI Governance architecture scaled into that environment because it was built on principles, not on the specific technical properties of any generation of AI.

The agentic era is the deepest extension yet. Agent cards are a first-class agentic AI Governance object in the Credo AI platform. Pre-deployment agentic AI Governance is a formal process — risk assessment, regulatory classification, business context configuration review, and sanctioning, all connected to the agent card and registry. Runtime agentic AI Governance validates agent behavior continuously against the agent card and the constitution. Verification is matched to the type of claim being checked, not collapsed into a single judge — and the tool gateway enforces the constitution and configuration deterministically at the point of action, before the action reaches the system that would carry it out. And because agentic AI Governance must operate at the speed of agents, Credo AI has deployed AI to do AI Governance work — knowledge discovery agents that continuously populate the registry, guardian agents that assess risk, retrieve evidence, and respond to incidents — with human oversight calibrated to the stakes of each decision.

Credo AI governs both agents enterprises build and agents they buy — including agents embedded in Salesforce, ServiceNow, Microsoft, and Workday. Credo AI has connected agentic AI Governance to MCP and the emerging interoperability protocols of the agentic ecosystem, so that agent-to-agent communication does not become an AI Governance blind spot.

AI Governance is the brain of the agentic AI stack. Security tools, observability platforms, and privacy systems are essential muscles. But muscles without a brain do not produce accountability. Building the AI Governance brain — with the registry, the agent cards, the constitutions with right verification methods, the pre-deployment gates, the runtime validation, the tool gateway, and the policy connection that makes it all meaningful — is what Credo AI has been building. And it is what the agentic era requires now.

[Learn More](#)

10

•

About the Authors

About the Authors



Navrina Singh

Founder & CEO

Credo AI

Navrina Singh is the Founder & CEO of Credo AI, the leading AI Governance and trust platform enabling enterprises and governments to scale AI with confidence, control, and accountability. She was named to the TIME100 AI 2025 list, recognizing the world's most influential leaders shaping the future of artificial intelligence.

Navrina served as a member of the U.S. Department of Commerce's National AI Advisory Committee, advising the White House on national AI strategy and Governance. She is also a UN AI Advisor and contributes to OECD-led global AI policy efforts, working at the intersection of AI, standards, guardrails and national competitiveness.

With over two decades of experience across enterprise AI and emerging technology Governance, Navrina advises Fortune 500 companies, boards, governments, and sovereign entities on AI strategy and operationalizing AI Governance—particularly for Generative and Agentic AI. She is a trusted advisor on aligning national AI ambition with durable Governance and long-term value creation, making AI trust measurable, defensible, and scalable.



[View on LinkedIn](#)

About the Authors



Ehrik Adalma

Head of Product

Credo AI

Ehrik Aldana leads product at Credo AI, where he translates the realities of AI policy and regulation into the Governance infrastructure enterprises can actually deploy. He joined Credo AI in 2022 and has been central to building its agentic AI Governance capabilities, including the Agent Registry. Ehrik is a Co-Founder and Board Member of The AI Education Project and previously worked with The Future Society and UC Hastings College of the Law. He holds a BA from Yale University.



[View on LinkedIn](#)

About the Authors



Ian Eisenberg

Leading AI Research

Credo AI

Ian Eisenberg leads Technical AI Governance Research at Credo AI, focused on system-level interventions that make AI more effective and more accountable. He combines deep technical expertise — as a computational social scientist and machine learning engineer — with a pragmatic view of how AI is actually developed and deployed. Ian is also the founder of the AI Salon, a forum for civil society dialogue on AI, and his work is motivated as much by AI's transformative potential as by the systemic risks that come with it. A cognitive neuroscientist and psychologist by training, he has been a researcher at Stanford, Columbia, Brown, and the National Institutes of Health.



[View on LinkedIn](#)

Contributors



Design and Layout

Lucy van der Meer

Brand and Design Manager

Credo AI

Lucy van der Meer leads brand and design at Credo AI, where she helps translate complex ideas around trusted AI Governance into visual experiences that leave their impact in the hearts of people.

Her background spans traditional arts, digital design, and everything in between. But what's always driven her isn't the medium — it's the mark. The moment a brand stops being something people see and starts being something they remember. Lucy works at the intersection of human psychology and design strategy, building brands that don't just look good, but mean something to communities all around the world.

Before Credo AI, Lucy launched and relaunched brands internationally, working across user research, web design, and brand identity to create experiences that stuck around long after the first impression. Like graffiti on a wall, weathered, real, and impossible to ignore.



[View on LinkedIn](#)

Contributors



Content Development

Jerome J. Sanders

Senior Content Marketing Manager

Credo AI

Jerome leads content marketing at Credo AI, where he helps Fortune 500 organizations build trusted AI Governance through data-driven strategic storytelling and thought leadership. With over a decade of experience in content strategy, product marketing, and product development at Cisco, he specializes in translating complex AI concepts into narratives that drive business outcomes, clarity, and industry influence for senior executives.

A 2024 Aspen Institute First Movers Fellow, United Nations Global Compact SDG Innovation Accelerator and Board Chair at Honor Your Dreams, Jerome focuses on expanding equitable access to foundational AI technologies. He holds a business degree from Santa Clara University's Leavey School of Business.



[View on LinkedIn](#)