



AI GOVERNANCE PLAYBOOK 2026

7 Building Blocks for Governing AI at Enterprise Scale

Beau Wyrick

AUTHOR

Stephanie Paradis

CO-AUTHOR



FIRST SAN FRANCISCO PARTNERS

Table of Contents

THE CASE FOR AI GOVERNANCE <i>Balancing Risk and Reward for Autonomous Decisioning</i>	Page 3
1. VALUE-FIRST USE CASES <i>Start with the decision, then prove its value</i>	Page 4
2. DECISION STEWARDSHIP <i>Accountability starts upstream of the model</i>	Page 4
3. AI-READY DATA AND THE SEMANTIC LAYER <i>Define meaning once, govern it everywhere</i>	Page 5
4. GOVERNING AI AGENTS <i>Every agent is an identity</i>	Page 6
5. SCOPE IS BINARY <i>Rigor is gradient</i>	Page 7
6. EXPLAINABLE AI <i>It's not what the AI says it is</i>	Page 8
7. LITERACY AND THE AI GOVERNANCE OPERATING MODEL <i>Governance is a job now, not a virtue</i>	Page 9

Balancing Risk and Reward for Autonomous Decisioning

A year ago, the case for AI governance was a case against recklessness. "Move fast and break things" was still the industry reflex, and the argument was about restraint.

That framing no longer holds, because neither side of the argument can win on its own terms. Move fast and break things, and you risk real harm, reputational damage and value destroyed inside your own business. Move too cautiously, and you never get started, never recoup the investment and get left behind. AI governance is the balancing mechanism between the two. No other focus area has the bandwidth, purpose or scope of decisioning — only AI governance. The only workable position sits between the two: AI governance that is equally oriented toward protecting the organization and producing value.

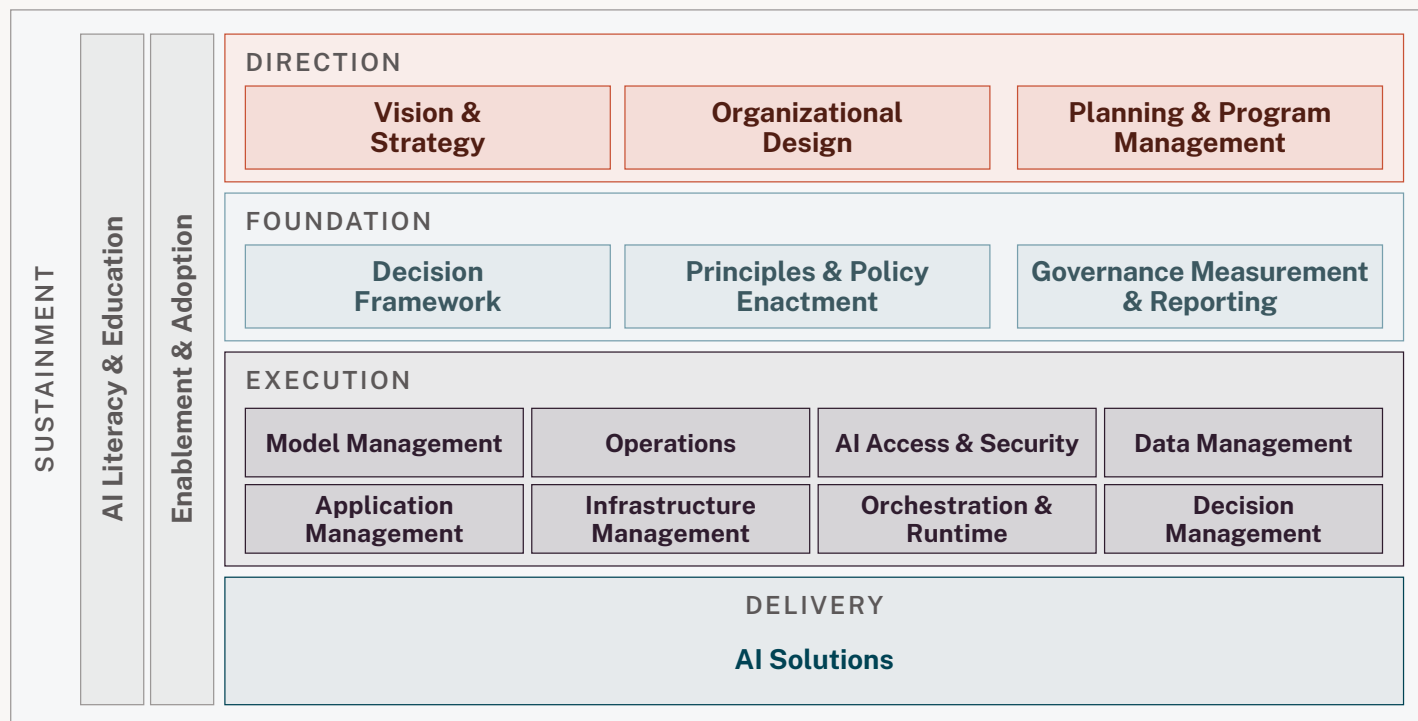
TWO THINGS CHANGED THE TERMS OF THE CONVERSATION

First, the board-level question is shifting. The conversation is moving away from what AI has returned in broad strokes and toward how the organization will build a sustainable competitive advantage around it, which pulls boards closer to shaping strategy rather than simply validating it.

Second, the thing being governed changed. Governance was built for solutions with relatively passive outputs, several steps removed from the business outcome and with human intervention built in. Now AI sits much closer to the outcome. A single individual can deploy tools that once required an entire department, and increasingly AI itself reasons, decides and acts, taking on decision authority previously reserved for people. Either way, the impact can ripple far beyond traditional boundaries.

This edition of the playbook reflects these shifts: seven building blocks, each updated for what has changed over the past year, and each with the questions every leadership team should be able to answer today.

AI Governance Framework



1 Value-First Use Cases — Starting with the decision balances risk and reward.

AI programs still fail for the reason they've always failed: **they launch without clarity on the specific decision AI is meant to support**, and on whether that decision is appropriate for a machine, or best validated by a human. That gap hides problems that surface downstream, including scope creep, an unclear impact radius and no agreed measure of success. Months later, the effort stalls or is scrapped, and the organization has lost more than money. Time, effort and employee morale all suffer.

DIRECTION

Vision & Strategy

Start instead with business process modeling. Understand, at the most granular level, the specific tasks AI is meant to augment, the decisions it will influence or make, and who is affected, directly and indirectly. Take that to the risk committee: **if the risk of the use case matches or exceeds its anticipated value, you've avoided sinking money into a lost cause.** That review is the Go or No-Go gate, the first Decision Stewardship control point covered in the next block. If the use case passes, define the financial measure, the baseline and the thresholds for scaling, pivoting or retiring it before the pilot starts, while it's still politically possible to shut it down. Then look across open use cases for a common core of shared data. Data work is the hard work, and reusing what's already done or underway is the key to rapid deployment.

WHAT'S CHANGED SINCE 2025

MEASURABLE VALUE / The bar moved from "aligned to strategy" to "provable at the P&L line." Research across multiple 2026 studies converges on the same uncomfortable finding: roughly 95% of enterprise AI pilots produced no measurable profit-and-loss impact.

GOVERNANCE, NOT MODELS / Gartner projects that 60% of AI projects will be abandoned through 2026 because organizations lack AI-ready data and coherent governance — meaning most failures aren't model failures at all.



ASK YOURSELF

For each active AI initiative, can we name the specific decision it supports, and whether a human or the AI acts on it? Are we comfortable with AI making and potentially acting on this decision and, if so, under what conditions? What business measure does it move, against what baseline? Do we have pre-agreed thresholds for scaling, pivoting or retiring it, and has anyone ever invoked them?

2 Decision Stewardship — Accountability starts upstream of the model.

Most AI governance starts too late. By the time there's a model to review, the highest-risk decisions have already been made: **how much control are we conceding to AI** in forming the output a critical business decision will rest on, and how much control are we conceding to AI to act on that decision? Decision Stewardship closes that gap.

EXECUTION

Decision Management

Decision stewardship is the ongoing, accountable management of AI-related decisions on behalf of the individuals and functions that bear their consequences. It is not a single approval, and it is not an audit of model outputs on a per interaction basis. It is the **discipline of ensuring that the point at which organizational authority converts into AI's authority to reason, decide, and act** is deliberately configured, continuously justified, and demonstrably trustworthy.

In practice, that accountability spans four lifecycle control points: Go or No-Go, before any resources are committed; Play, Pause or Rewind, once the data and the problem are better understood; Risk-Reward Evaluation, before the solution goes live; and Monitor and Adjust or Retire, for as long as it runs. Three of the four sit outside the build itself, which is exactly why a single pre-launch review isn't enough.

2

Decision Stewardship, cont. — Accountability starts upstream of the model.**WHAT'S CHANGED SINCE 2025**

RISK ISN'T TITLED / Org-chart seniority is no longer a usable proxy for stakes. A low-cost tool touching medium-to high-risk uses such as hiring, lending or clinical triage can carry enormous consequence without ever reaching an executive's desk.

APPROVAL ISN'T ENOUGH / Autonomy broke the one-time-approval model. An agent that was well-governed at launch can drift out of alignment with the problem it was built to solve, which makes ongoing decision fitness a governance requirement rather than an audit finding.

IMPACT OUTSCALES AUTHORITY / AI is decoupling the scope of impact from organizational position. Individuals now have tools that execute at a scale and speed that once required the resources of entire departments, and the impact of their output can ripple far beyond their traditional spans of authority.

**ASK YOURSELF**

For our highest-consequence AI systems, can we name the individual accountable at each of the four lifecycle control points? Have we identified all of the inputs, influencing factors, and outcome paths for a given use case decision? If a decision an AI system influenced were challenged tomorrow, who answers? And with what evidence?

3

AI-Ready Data & The Semantic Layer — Landing and expanding consistency.

The BI era left a lesson most organizations are still cleaning up: **when business meaning is tightly coupled with a specific application, every tool ends up with its own version of the truth.** Reporting could survive that, because it simply returned values that a human ultimately interpreted and acted on. AI can't. It is a dynamic system that acts on data, applying inference at every interaction to arrive at a decision or output. It infers meaning, not just values, and an agent that can read your data and your prompt but can't interpret your business will produce confident, eloquent, and potentially wrong answers.

That is why metadata has moved from nice-to-have reference material to operational control. Definitions, relationships, policies and lineage are the translation key AI runs on, and the semantic layer is where they become operational, adding judgment and governance on top of definitions and logic. Componentized AI strategies capitalize on the value of a **universal semantic layer: a technology-agnostic layer between your data platforms and the tools and agents that consume them**, where meaning is defined once, propagated across the enterprise and updated in one place. It also gives AI a governed path to the right data instead of leaving a model to guess at queries. Any effort towards building a semantic layer strengthens traditional data governance along the way, because data can only really be governed through its metadata.

Be selective. Many vendors now market a semantic layer, but not all of them enable portability on the outputs. The test is whether it stays agnostic to the stack underneath and covers tools outside of that vendor's offerings, or whether it offers partial coverage that ties you more tightly to one vendor's technology and roadmap.

EXECUTION**Data Management**

3

AI-Ready Data & The Semantic Layer, cont. — Landing and expanding consistency.**WHAT'S CHANGED SINCE 2025**

AMBITION OUTRUNS READINESS / Confidence hasn't caught up with ambition: only 37% of organizations report confidence in their data practices, even as AI investment accelerates.

METADATA BECAME OPERATIONAL / Metadata shifted from documentation to enforcement. Definitions, policies and quality signals are increasingly expected to flow into the systems where work happens, and into the context AI systems consume at runtime.

CONTEXT IS INFRASTRUCTURE / Business context moved from a nice-to-have to infrastructure. At its 2026 summit, Gartner framed context as "the new critical infrastructure," on the grounds that agents cannot operate reliably without shared business meaning that goes beyond raw data.

**ASK YOURSELF**

Have we documented meaning and context sufficiently for AI to unambiguously reason over? Is there any tacit knowledge or assumptions applied in this decision we need to explicitly capture? If we're evaluating a semantic layer, does it cover our whole stack, or just one vendor's corner of it?

4

Governing AI Agents — Every agent is an identity.

As the reach and authority of AI agents expands, so do **concerns over visibility into what it is allowed to do — and what it could do outside of that boundary**. While agents introduce a new class of identity into the business, the good news is every identity, human or not, is governed using the same principles of permissions, roles and scopes. Agentic identity management doesn't start at zero — it extends what already exists.

DELIVERY**AI Solutions**

Anxiety over agentic identity includes, in part, a subconscious recognition of two attributes "baked" into existing identity management strategies we rely on that don't translate well to AI — human judgement or self-restraint on what is acceptable behavior and explicit failure. Today's probabilistic models are built to generate statistically likely answers, not permissioned ones. Therefore, there's no guarantee an agent will respect an instruction boundary the same way every time and that a violation will be noticed. So, what used to happen instinctively now has to be engineered on purpose. That's where governance stops being a policy exercise and becomes a business one: **a chosen use case determines the AI architecture, and that architecture determines exactly what metadata, permissions, and controls are needed to keep it operating as intended**. Governance coordinates and facilitates these conversations and then continuously manages the metadata, which in turn, become operational controls for AI.

WHAT'S CHANGED SINCE 2025

THE AGENT EXPLOSION / The volume problem became real. Gartner projects that the average global Fortune 500 enterprise will have more than 150,000 agents in use by 2028, up from fewer than 15 in 2025.

POLICY LAGS PRACTICE / Intent is far ahead of practice. Only 44% of organizations have implemented any policy to manage AI agents, even though 92% agree that governing them is critical.

4

Governing AI Agents, cont. — Every agent is an identity.**WHAT'S CHANGED SINCE 2025**

SHADOW AI IS NOW AN ACCESS ISSUE / Shadow AI stopped being a policy violation and became an access-path problem. Verizon's 2026 Data Breach Investigations Report found frequent employee use of unapproved AI tools tripled to 45% of the workforce in a single year, and Saviynt reported that 75% of CISOs have already found unsanctioned AI tools running in production.

MACHINES GETTING TOO MUCH ACCESS / Access is being granted more freely to machines than to people. Roughly 70% of organizations give AI systems more access than they would give a human employee doing the identical job.

**ASK YOURSELF**

What is this agent actually allowed to do and did we decide that on purpose, or did it just end up that way? Where is that boundary documented and could we point to it if asked? If this agent did something outside its boundary, is there a visible sign or would it just happen quietly?

5

Scope Is Binary — Rigor is gradient.

The instinct with a new risk is to write one all-encompassing policy and apply it everywhere. With AI, that instinct backfires in both directions: it strangles low-risk work and under-protects the systems that matter. The fix is to separate two questions most organizations blur together. **Scope asks whether a system falls under AI governance at all, and the answer is binary. Rigor asks how much governance it needs, and the answer scales with risk.**

FOUNDATION**Principles & Policy
Enactment**

Set scope with a single gate: are any of the system's parameters or logic inferred from data, rather than fully specified by a person? If so, it's in scope, however simple, familiar or long-standing it is. A transparent three-coefficient regression carries the same kind of provenance risk as a neural network; it's just easier to validate. Apply the gate system by system, not team by team, because a single workflow often mixes hand-authored rules and data-inferred models. Fully rule-based systems stay within "traditional" analytics governance, unless their rules are quietly regenerated by a recurring data-driven process.

Then let rigor do the proportional work. **Tier each in-scope system on factors such as consequence, autonomy, opacity and compliance obligations**, using established standards like NIST AI 100-1 rather than inventing your own taxonomy. A churn model that has run for years stays in scope but lands in a lower tier; an agent acting on customer accounts lands high. That is what makes minimum viable governance possible: nothing slips out of view, low-risk systems aren't buried in process, and the business is far more likely to adopt governance it doesn't experience as a blanket tax. At the higher tiers, push controls into runtime; monitoring of performance should be at a more granular level, occur more frequently, and require higher decision-level authority.

WHAT'S CHANGED SINCE 2025

BINARY GOVERNANCE BREAKS DOWN / All-or-nothing rigor now has a documented failure rate. Gartner predicts that by 2027, 40% of enterprises will demote or decommission autonomous AI agents because of governance gaps found only after a production incident — and names binary governance, locked down or fully trusted, as the root cause.

CONTROLS MOVED TO RUNTIME / Enforcement is moving into the runtime layer. Gartner published its first Market Guide for Guardian Agents in February 2026, defining them as a blend of AI governance and runtime controls that supervise agent behavior and intervene when it deviates.

5 Scope Is Binary, cont. — Rigor is gradient.

WHAT'S CHANGED SINCE 2025

CONSEQUENCES GOT REAL / Consequences became literal. Gartner expects legal claims tied to AI-related safety failures to exceed 2,000 by the end of 2026, driven by insufficient risk guardrails.



ASK YOURSELF

Do we have a single, written test for what falls under AI governance, or does it depend on which team built the system? Does the level of control actually differ by risk tier? Can we use existing, well-understood systems as a starting indicator around which to build our risk profile? Which of our guardrails execute automatically at runtime, and which rely on someone remembering to check?

6 Explainable AI — It's not what the AI says it is.

When an AI system makes or shapes a decision, "the model did it" is not a defense. **Courts are already treating AI tools as extensions of the organizations that deploy them**, and a vendor's model doesn't shield the business that chose to rely on it. Accountability for the outcome stays with you, whether or not you can see inside the system.

EXECUTION

Model Management

Often, you can't. Most modern AI is probabilistic, and no one will ever fully account for every parameter inside a neural network. Asking the system to explain itself doesn't close the gap either: it produces a plausible reconstruction, not a verified account of the computation behind the result. **What you can govern is everything around the model**, including the sources it may draw on, the data fields it may use, the version in production, the approved reason codes behind each decision and the human accountable for review. Document those before the build as requirements, and they become your evidence afterward.

Use the NIST AI Risk Management Framework to **define what trustworthy AI means for your organization**, and the characteristics your systems must demonstrate to be responsible, defensible and trusted. Then, identify what metrics would indicate that the system is behaving according to those principles; these metrics will translate system behavior to interaction outcomes. For example, Context Grounded Accuracy Rate is part of the picture that defends whether the AI is producing valid and reliable answers.

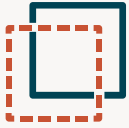
WHAT'S CHANGED SINCE 2025

VENDOR SHIELD IS GONE / Courts stopped accepting "the vendor built it." In *Mobley v. Workday*, a federal court allowed discrimination claims to proceed on the theory that an AI screening tool can act as an agent of the employer using it.

COMPLIANCE DIDN'T PAUSE / The regulatory clock kept running. The EU's Digital Omnibus on AI deferred most high-risk obligations to December 2027 and August 2028, but Article 50 transparency obligations apply from 2 August 2026, and the Article 4 AI literacy duty was left untouched.

DISCOVERY DEMANDS EVIDENCE / Opaque decision logic is heading into discovery. In February 2026, the Fourth Circuit revived class claims alleging that Navy Federal's mortgage underwriting algorithm relied on variables and weights kept undisclosed, even from denied applicants. Liability hasn't been decided as of publication of this playbook, but discovery will demand a defensible account of what drove each decision.

6

Explainable AI, cont. — It's not what the AI says it is.**ASK YOURSELF**

For our highest-consequence AI decisions, could we explain what drove a specific outcome without asking the model? Do we document sources, data fields, model version and reason codes before the build, or reconstruct them after a challenge? Have we operationalized our metadata to support our decisioning guardrails? Have we identified the metadata to uphold the desired principles and enact appropriate controls? If a vendor's model made the call, are we prepared to own it?

7

Literacy & The AI Governance Operating Model — Governance is a job now, not a virtue.

The best model won't drive impact if your workforce can't engage with it, apply its outputs appropriately or recognize its limits. But literacy alone isn't a program, and one-size training rarely changes how people work. **People learn far more from experience than from mandate**, and no governance team can build tailored programs for every role at scale.

SUSTAINMENT**AI Literacy & Education**

Someone has to own AI governance, with a budget, a mandate and a place in the operating model. Then let the operating model do the teaching. **Form a pod around each AI solution with six roles:** an AI Engineer who builds it, an AI Steward who validates that it works and is operationally sound, a Decision Steward who owns the risk and value of the decision it supports, a Data Steward who ensures the data and metadata foundation is governed, a Business SME who brings domain judgment, and an AI Communications lead who drives adoption and literacy. Pods persist until the solution is retired, and each member is accountable to both the pod and their governing committee. That is how minimum viable governance gets activated while real solutions get built.

The pod complements formal training by making it relevant. Members learn from each other, toward tangible goals, how architecture, deployment context, data and business decisions shape one another. The result is internalized knowledge of where AI's boundaries are, where human judgment is needed in ambiguous situations and how much skepticism its outputs deserve. Pods also open channels that surface existing solutions, curb agent proliferation and keep teams from relitigating settled decisions, so insight flows across the organization instead of staying trapped inside a single implementation.

WHAT'S CHANGED SINCE 2025

LITERACY IS NOW REQUIRED / AI literacy became an obligation rather than an aspiration. The Article 4 literacy duty under the EU AI Act was untouched by the 2026 deferrals.

CROSS-FUNCTIONAL IS NO LONGER OPTIONAL / The cross-functional argument is over. The open question is no longer whether AI needs perspectives beyond the technology team — it's who holds the decision rights when those perspectives disagree.

THE FUNCTION IS FORMALIZING / The function is formalizing. Gartner expects AI governance programs with dedicated headcount and specialized software to become the norm, driven by both regulatory and business requirements, and managed independently of security.

**ASK YOURSELF**

Who owns AI governance at our organization by name, and what authority do they actually hold? Does each AI solution have a standing, cross-functional team accountable for it from build through retirement? Are people learning to work with AI by doing it, or did we deliver one training to everyone and call it coverage?

Partner with FSFP

At First San Francisco Partners (FSFP), we help enterprise organizations design and implement AI governance strategies that unlock innovation while ensuring compliance and alignment with business outcomes.

With nearly two decades of experience turning data into a strategic asset, our senior consultants partner with you to build and operationalize the foundational capabilities: data governance, metadata management, semantic clarity and decision accountability, that make responsible AI governance possible at scale. These are the cornerstones that separate AI programs that prove their value from AI programs that get quietly retired.

READY TO TAKE THE NEXT STEP?

Start with a clear-eyed read on where you stand. Our AI Readiness Assessment takes a few minutes and tells you which of these seven building blocks needs attention first.

[Take the AI Readiness Assessment →](#)