



TURNING AI RISK INTO ADVANTAGE

AI GOVERNANCE · WHITEPAPER · SEPTEMBER 2026

From AI Governance Frameworks to AI Control Infrastructure

Most institutions don't need another AI governance framework. They need to turn the one they already have into an operating control system.

The 2026 risk environment, why the traditional operating model breaks, and a sequenced pathway for banks, fintechs and other regulated institutions.

Satyaki Ghosh Dastidar
AI Risk Consulting

At a glance

Most institutions do not need another AI governance framework. They need to turn the framework they already have into an operating control system.

THE ARGUMENT IN ONE PARAGRAPH

AI governance gets designed as policy, committees and standards, and almost never as the control infrastructure those things depend on. The half that would carry it to a running system — registration, classification, control provisioning, evidence — is rarely scoped, budgeted or owned. Institutions therefore govern a shrinking share of their own AI estate, and cannot measure by how much.

18%

of US firms had adopted AI by December 2025

Census BTOS, via the Federal Reserve

78%

of the labor force works at a firm that has adopted AI

Survey of Business Uncertainty, Nov 2025

80m+

fraud signals a day through one bank's agentic system

FSB consultation report, June 2026

Five things to take away

- 1 The unit of risk moved from the model to the system.** When a system can act, the question is authority, not accuracy — and a control built to review an output cannot govern a decision loop.
- 2 The supervisory perimeter narrowed while the risk widened.** SR 26-2 places generative and agentic AI outside its scope and directs institutions to their own practices instead — a transfer of responsibility, not a reduction.
- 3 Governance capacity is itself a risk factor.** Leakage to the ungoverned path begins well before saturation, and every throughput metric improves while it happens.
- 4 Third-party dependency is structural.** The object to be governed is no longer the model. It is the dependency graph around it.
- 5 The fix is architectural, not editorial.** Registration wired into deployment, classification computed from attributes, controls provisioned from it, evidence emitted by the running system.

WHAT TO DO FIRST

Secure the chokepoint. Identify the small number of paths through which AI reaches production — pipelines, model-provider credentials, procurement, the assistant platform — and make them the mandatory route. Every control in this paper depends on the system being observable.

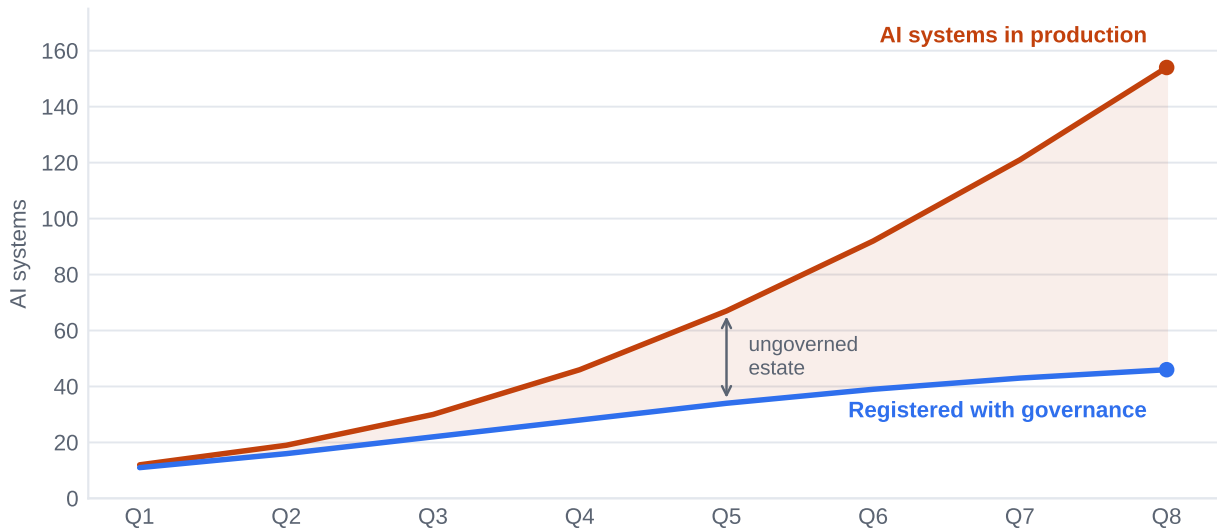
1 The 2026 AI risk environment

Adoption is broader than firm-level counts suggest

The Federal Reserve Board's April 2026 analysis of AI adoption draws on several surveys that appear to disagree and in fact do not. The Census Bureau's Business Trends and Outlook Survey put firm-level adoption at 18% of US firms as of December 2025. The Survey of Business Uncertainty, which weights by employment rather than by firm, found 78% of the labor force working at firms that had adopted AI and 54% at firms using large language models. A third measure found 41% of the workforce using generative AI at work.

The divergence is methodological. Adoption is concentrated in larger firms, so counting firms understates exposure while counting workers overstates the number of adopting organizations. For a risk function the employment-weighted figure is the relevant one: in an institution of any scale, the working assumption should be that AI is already in use across the enterprise, whether or not it has been registered.

EXHIBIT 1 Registration does not keep pace with deployment



Illustrative — the shape of the dynamic, not measured values.

The governed population grows linearly with review capacity. The deployed population does not. The gap between the two lines is the part of the estate no control touches.

Deployment in financial services is already operating at scale

The Financial Stability Board's June 2026 consultation report on sound practices documents institutions using AI across fraud detection, anti-money-laundering, customer service, risk management and operations. Its most instructive example is a large, internationally active bank operating an agentic fraud system that monitored more than 80 million signals each day across transactions, card and online payments and digital banking interactions. The agent contributed to developing or updating three quarters of the bank's card fraud rules, and the deployment helped reduce fraud losses by over 20% in the first half of the 2026 financial year against the same period in 2025.

Two features of that example matter more than the headline benefit. A system generating rule changes continuously cannot be assured by a control that operates annually. And the bank retained human review and approval of every new rule before implementation — which is to say the value was realized by instrumenting the system, not by exempting it.

Concentration is a recognized vulnerability

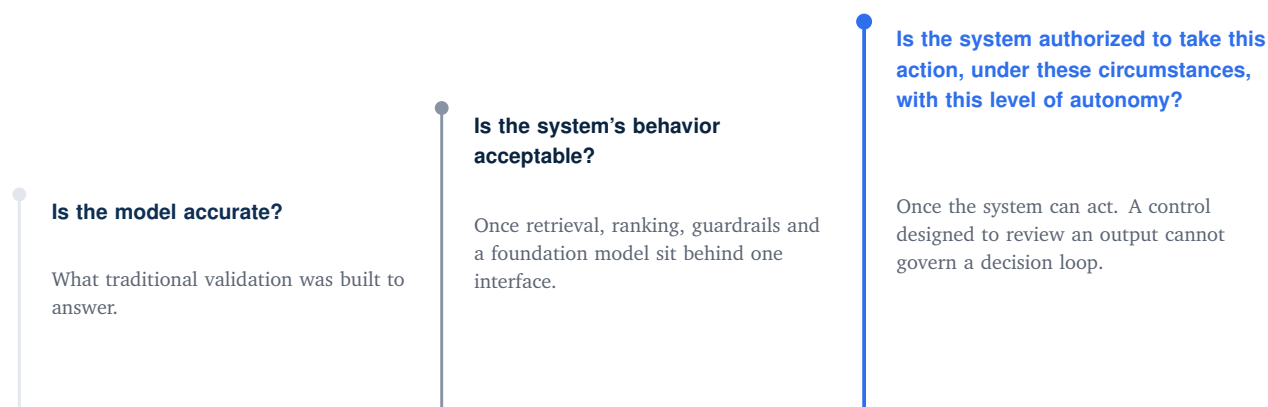
Both the FSB and the BIS have identified third-party dependency and provider concentration as material AI-related vulnerabilities. The FSB observes that providers of cloud infrastructure, hardware and AI foundation models are often highly concentrated and in some cases vertically integrated, with geographic concentration at key points in the supply chain, and warns that reliance on common models, datasets or infrastructure can produce correlated behavior across institutions.

That is a systemic framing of what is, inside a single institution, an ordinary operational question that most cannot answer quickly: which of our production systems depend on this provider, and what happens if it changes or stops?

2 Why the traditional operating model breaks

The model risk operating model most institutions run was designed against a coherent set of assumptions. A model is a discrete artifact. It is built in-house or licensed and inspectable. Its behavior is stable between releases. Its output is a number a human then acts on. And the population grows slowly enough that a central function can review each one.

Every one of those assumptions is now contingent. Traditional model governance asks whether a model performs acceptably. That question becomes insufficient the moment the model is one component inside a composed system — and insufficient in a different way again once that system can act.

EXHIBIT 2 The question governance has to answer keeps escalating

Each step subsumes the one before it. The third is not a harder version of the first — it is a different control problem, and periodic review of outputs does not address it.

EXHIBIT 3 The operating model was built for different objects

Dimension	Traditional MRM	What AI governance requires
Object governed	The model	The system: model, retrieval, tools, guardrails, humans
Behavior	Stable between releases	Can change without a release
Output	Deterministic given inputs	Probabilistic and path-dependent
Assurance	Periodic validation	Continuous assurance
Provenance	Internal or inspectable	Model ecosystem, largely external
Decision locus	Human decides	Human, model and agent share the decision
Register	Model inventory	Dependency graph
Classification	Committee judgment	Computed from attributes
Evidence	Validation artifacts	Runtime telemetry

The left column is the set of assumptions the operating model was built on. The right column is what holds instead.

The failure mode is not that the framework is wrong. An annual validation cadence against a system whose behavior can change without a release is not a weak control; it is a control measuring the wrong thing at the wrong frequency.

3 The regulatory boundary

SR 26-2 and a deliberate exclusion

SR 26-2 was issued on 17 April 2026 by the Federal Reserve Board, the OCC and the FDIC, applying to banking organizations with more than \$30 billion in total assets. It narrows the definition of a model, introduces materiality-based tiering, and — most consequentially for this paper — places the fastest-growing part of the AI estate outside its own perimeter.

THE SCOPE LANGUAGE, VERBATIM

“Generative AI and agentic AI models are novel and rapidly evolving. As such, they are not within the scope of this guidance.”

And for what falls outside: “A banking organization’s risk management and governance practices should guide the determination of appropriate governance and controls for any tools, processes, or systems not covered in this document.”

EXHIBIT 4 What the guidance covers, and what it hands back to you

In scope of SR 26-2	Governed by your internal standard alone
Traditional statistical and quantitative models	Generative AI, in every use case
Non-generative, non-agentic AI models	Agentic systems that take action
Institutions above \$30bn in assets	Everything at institutions below the threshold
Complex quantitative methods producing estimates	Deterministic rules and simple calculations

The right-hand column is not unregulated. It is regulated by an internal standard that supervisors will still expect to see applied and evidenced.

The boundary is deliberate and temporary — an acknowledgment that prescriptive guidance would date quickly, not a judgment that the risk is low. Two errors follow from misreading it. The first is treating out-of-scope as out-of-governance. The second is the opposite: extending the full model risk apparatus to every generative use case, which produces the capacity dynamics of Section 4.

The wider supervisory context

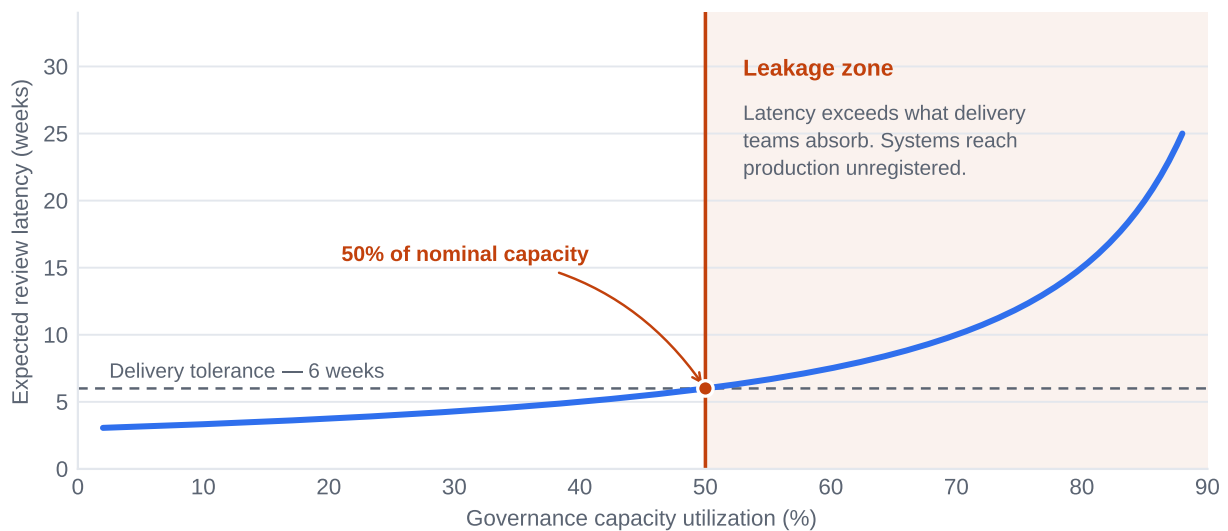
The August 2026 reform of the unsafe-or-unsound practice and MRA standards, adopted by the OCC and the FDIC but not the Board, redefines supervisory findings around likely material financial harm and demotes process and documentation weaknesses to supervisory observations. Read alongside SR 26-2, several of the routes by which an AI governance weakness would previously have generated a supervisory finding no longer reliably do so. Responsibility does not disappear. It relocates to the board and to the institution’s own control architecture.

4 The capacity problem

Governance throughput is not an efficiency concern. It is a control failure with a misleading signature, and it is the reason coverage erodes without anyone deciding that it should.

The traditional operating model assumes the number of systems entering governance is manageable relative to the number of people reviewing them. Under accelerating adoption that assumption fails — and it fails in a way that looks like success.

EXHIBIT 5 Review latency does not degrade gracefully



Single-server queue, three-week mean review time. Derivation in Appendix A.

Teams do not defect when the queue saturates. They defect when expected latency exceeds what their delivery cycle can absorb — which, for a three-week review against a six-week tolerance, happens at half of nominal capacity.

Defection removes demand from the queue. Registered arrivals fall, backlog stabilizes, time-to-approval improves, and reviewer utilization looks comfortable. Every operational metric the governance function reports gets better. Meanwhile the deployed population keeps growing and governance coverage falls.

THE FAILURE SIGNATURE

Backlog flat or falling, time-to-approval improving, reviewer utilization comfortable — while the number of AI systems in production grows faster than the number of registrations. Coverage is the only metric that reports the truth, and almost no institution measures it.

The queue is self-stabilizing, and it stabilizes by leaking.

Coverage cannot be measured from inside the governance process, because the systems that matter are by definition not in it. It has to be triangulated from outside: procurement and vendor spend, cloud and API egress to known model providers, code repositories and CI pipelines, identity and service-account activity. Reconcile that population against the register; the gap is the number.

The two levers are therefore reducing review time so that defection is irrational, and closing the discovery gap so that defection is visible when it happens. The sections that follow attack both.

5 The AI control plane

The organizing idea of this paper is that governance should be implemented as a control plane — a set of connected, largely automated stages every AI system passes through and continues to pass through — rather than a sequence of documents produced at intervals.

EXHIBIT 6 The control plane, and where the loop closes



The stage names are largely familiar. What differs is that each is triggered by a system rather than by a meeting, and that the loop returns to classification whenever the system changes.

The difference from the traditional lifecycle is three properties of the connections, not the stages themselves.

- 1 **The stages are wired to systems, not to meetings.** Registration is triggered by the deployment pipeline. Classification is computed from registered attributes. Controls are provisioned from the classification. Evidence is emitted by the running system.
- 2 **The loop closes.** Reclassification is an ordinary event rather than an exception. A system that acquires a new tool, data source, user population or degree of autonomy re-enters classification automatically and may acquire controls it did not previously have.
- 3 **Human judgment is spent where it is scarce.** Committee time goes to setting the rules, adjudicating genuine ambiguity and reviewing the highest consequence tier — not to processing volume.

Only half the system was ever designed — the half that gets written down, and not the half that would carry it to a running system.

6 The operational inventory

The highest-leverage change available to most institutions is to stop treating the inventory as a periodic data-collection exercise. A register populated by asking teams what they have built is stale on the day it is compiled and silent about everything that never answered.

EXHIBIT 7 What an operational inventory holds that a model register does not

Field	Why it is load-bearing
Purpose and decision surface	Which decisions the system influences or takes
Ownership	An accountable individual in the first line, not a team mailbox
Data	Categories consumed, retained and emitted, with sensitivity
Model and vendor dependencies	Provider, model family, and the version or endpoint in production
Reachable systems	What it can call, read or write — the practical definition of its authority
Autonomy	Whether it recommends, acts with confirmation, or acts unattended
Version state	What is running now, and what changed at the last change
Dependency failure behavior	What happens if a provider changes the model or becomes unavailable

The last three are the fields traditional model inventories almost never hold, and the ones the 2026 environment makes indispensable.

THE INVENTORY TEST

Can you determine, in minutes, which production systems depend on a particular AI provider — and what would be affected if that provider became unavailable tomorrow? If the answer requires a survey, you have a register. You do not yet have an operational inventory.

7 Computed risk classification

Classification should be derived from recorded attributes rather than negotiated per deployment. The rubric below scores eight attributes on a three-point scale, each read from the inventory record. It is deliberately coarse: the purpose is consistent, reproducible triage at volume, not precision.

EXHIBIT 8 An attribute rubric for design-time classification

Attribute	1 — Low	2 — Elevated	3 — High
Consequence	Internal productivity	Customer or financial impact, recoverable	Material financial, safety or regulatory harm
Autonomy	Recommends only	Acts with confirmation	Acts unattended
Decision authority	Informs a human decision	Makes a reversible decision	Makes or executes a binding decision
Data sensitivity	Public or internal	Confidential or personal	Regulated or material non-public
External exposure	Internal users only	Customer-visible output	Direct customer or public interaction
Third-party dependency	Fully internal	External model, pinned version	External model, provider-managed updates
Reversibility	Trivially reversible	Reversible with effort	Irreversible or externally visible
Concentration	Provider used by few systems	Shared across a business line	Shared across critical processes

BASELINE

ELEVATED

TOP TIER

highest attribute is 1

highest attribute is 2

any attribute at 3

Tier is set by the highest attribute score, never the average. One attribute at 3 is sufficient on its own.

Why the highest score governs, not the average

This is the single most important design choice in the rubric. Averaging permits a system with unattended authority over irreversible actions to be pulled down into a light tier by six benign attributes — precisely the failure the classification exists to prevent. Concretely: any attribute at 3 places the system in the top tier; otherwise any attribute at 2 places it in the middle tier; everything else is baseline. The three bands are exhaustive and mutually exclusive by construction.

Two properties that matter more than the rubric itself

It is recomputable. Because the inputs are inventory attributes, the classification updates when the system changes. Granting an assistant write access to a system of record changes its reachable-systems and reversibility scores and re-tiers it automatically, without anyone remembering to ask.

It is auditable. The tier is a function of recorded facts, so “why is this system in this tier?” has an answer that does not depend on who was in the room. Where judgment overrides the computed tier, the override and its rationale are recorded as an exception — which makes the exercise of discretion visible and reviewable rather than invisible and routine.

8 Runtime assurance

Point-in-time validation establishes that a system was acceptable when someone reviewed it. Runtime assurance establishes that it still is. What to instrument differs by system type.

EXHIBIT 9 What runtime evidence has to establish, by system type

System type	What the evidence has to establish
Generative	Output conformance to policy; refusal and guardrail activation rates; drift in response characteristics; human override and correction rates
Retrieval-augmented	Groundedness of claims in retrieved sources; retrieval quality and coverage; staleness and permission scope of the corpus; behavior when retrieval returns nothing
Agentic	Action-level authorization; tool-call patterns against the permitted set; trajectory length and loop detection; intervention and rollback events; consequence of completed actions, not merely outputs
Multi-model	Which model served which request; routing and fallback behavior; consistency across components; end-to-end rather than per-component performance
Vendor models	Version and endpoint in use at time of request; provider-side change events; behavioral drift across an unchanged application; availability and degradation
Human-in-the-loop	Override rate, direction and latency; automation-bias indicators such as approval time collapsing toward zero; escalation completion

Instrument in proportion to consequence. The classification in Exhibit 8 is what makes that decision defensible.

Two cautions. The objective is not to instrument everything equally, but to generate continuous evidence where the consequence justifies the cost. And human-in-the-loop is not itself a control unless it is instrumented: an approval step whose median completion time has fallen to two seconds is a record of consent, not an act of oversight, and measuring it is the only way to know which one you have.

9 Implementation sequence

The sequence matters more than the pace. Each step is chosen because it makes the next one cheap; institutions that attempt them in a different order generally stall at step four.

EXHIBIT 10 A sequenced pathway, and what each step unlocks

STEPS 01–03 Buy back capacity

01 Secure the chokepoint

Make the small number of paths to production the mandatory route. Every later control depends on the system being observable.

02 Automate registration

Registration becomes a pipeline step, not a form. Compliance cost approaches zero; avoidance cost becomes not shipping.

03 Deterministic classification

The majority of use cases classify without consuming committee time — the fastest available reduction in review latency.

STEPS 04–06 Buy back assurance

04 Instrument the top tier

Runtime evidence is expensive to build and cheap to extend. Start where consequence is highest and the population smallest.

05 Automate evidence

Activations, overrides, escalations, tool calls, version changes and exceptions flow to one evidentiary store by default.

06 Close the loop

Thresholds trigger escalation and reclassification without a scheduled review. Periodic work becomes exception handling.

Doing the second band before the first is the common failure: assurance built on an estate you cannot see only assures the part that was already visible.

THE MEASURE OF PROGRESS

Not the number of systems reviewed. Governance coverage — registered systems as a share of systems actually in production — and mean time from submission to decision. The first tells you whether you are governing your estate. The second tells you whether you will still be governing it next year.

10 The board agenda

Five questions distinguish an institution that has AI governance from one that has AI governance documents. None can be answered from a policy.

- 1 **What is our governance coverage, and what is the trend?** How many AI systems are running in production, how many are registered, and is the gap widening or closing?
- 2 **Which of our AI systems can take action without a human, and what can they reach?** If this requires research to answer, the inventory is not operational.
- 3 **What is our largest AI provider dependency, and what is our exposure if it changes or fails?** Including the systems that depend on it indirectly.
- 4 **How long does the governed path take, and how long does the ungoverned path take?** The ratio predicts the coverage trend better than any control assessment.
- 5 **For our highest-consequence AI systems, what evidence do we hold that they are behaving acceptably today — not at last review?**

The board's position deserves particular attention in the current environment. With generative and agentic AI outside the scope of SR 26-2, and with process and documentation findings demoted under the August 2026 reforms, the external validation boards have historically relied on to confirm that AI governance is adequate is materially thinner than it was. The board is increasingly the party that has to satisfy itself, on evidence, that the control architecture works.

11 Conclusion

The frameworks are not the constraint. Most institutions have written down a defensible account of how AI should be governed. The constraint is that the account describes activities performed by people at intervals, against an estate growing faster than those people can review and changing faster than those intervals can detect.

Moving from governance as documentation to governance as infrastructure is not a technology project layered on the existing framework. It is the operationalization of the framework that already exists: registration wired into deployment, classification computed from attributes, controls provisioned from classification, evidence generated by the running system. Human expertise is not displaced. It is relocated to the ambiguity, the exceptions and the consequential decisions — the work it was always the scarce resource for.

THE THESIS IN ONE LINE

The safest AI governance model is the one in which the governed path is also the fastest path. Every institution that fails to achieve that will govern a shrinking share of its own AI estate — and will not be able to tell.

A Appendix A The capacity model

The claim in Section 4 — that leakage to the ungoverned path typically begins at roughly half of nominal review capacity — follows from a standard single-server queue. It is included here for readers who want the derivation; nothing in the argument depends on the reader accepting the model's simplifying assumptions.

Let λ be the rate at which new AI use cases enter governance and μ the rate at which reviews complete, where $\mu = R \times c$ for R reviewer FTE at c completed reviews per FTE per period. Utilization is $\rho = \lambda/\mu$. Expected time in system is

$$W = \frac{1}{\mu - \lambda} = \frac{t_s}{1 - \rho}, \quad t_s = \frac{1}{\mu}$$

where t_s is mean service time. Latency is not linear in load: at $\rho = 0.5$ it is twice service time; at 0.8, five times; at 0.9, ten times; at 0.95, twenty times.

Teams do not defect when the queue saturates. They defect when expected latency exceeds the delay their delivery cycle can absorb. Letting T be that tolerance, defection begins when $W > T$, giving a critical utilization

$$\rho^* = 1 - \frac{t_s}{T}.$$

With a three-week mean review against a six-week tolerance, $\rho^* = 0.5$. Halving mean review time to 1.5 weeks moves ρ^* to 0.75 and roughly triples the governable volume $\lambda^* = \rho^* \mu$, from 0.17 to 0.50 use cases per week. Note that under this single-server model, adding reviewers and shortening review are the same lever: both reduce t_s . Distinguishing them requires a multi-server treatment this appendix does not attempt.

Defection then removes demand from the queue. Observed utilization $\rho_{\text{obs}} = \lambda_{\text{registered}}/\mu$ stabilizes or declines while true demand $\lambda_{\text{true}} = \lambda_{\text{registered}} + \lambda_{\text{shadow}}$ continues to rise, so governance coverage

$$C = \frac{\lambda_{\text{registered}}}{\lambda_{\text{true}}}$$

falls while every throughput metric improves. This is the failure signature described in Section 4.

SOURCES

Board of Governors of the Federal Reserve System, FEDS Note, *Monitoring AI Adoption in the U.S. Economy* (3 April 2026), drawing on the Census Bureau Business Trends and Outlook Survey, the Survey of Business Uncertainty and the Real-Time Population Survey. Financial Stability Board, *Sound Practices for the Responsible Adoption of Artificial Intelligence*, consultation report (June 2026). Financial Stability Board, *Monitoring Adoption of Artificial Intelligence and Related Vulnerabilities in the Financial Sector* (October 2025). Board of Governors, OCC and FDIC, SR 26-2 / OCC 2026-13, *Revised Guidance on Model Risk Management* (17 April 2026). OCC and FDIC final rule, *Unsafe or Unsound Practices, Matters Requiring Attention* (August 2026); OCC PPM 5310-3 and PPM 5400-11. Bank for International Settlements, Financial Stability Institute, supervisory insights on AI in financial services.



TURNING AI RISK INTO ADVANTAGE

The safest AI governance model is the one in which the governed path is also the fastest path.

ABOUT AI RISK CONSULTING

AI Risk Consulting advises banks, fintechs and other regulated institutions on AI governance, model risk management and runtime reliability. Satyaki Ghosh Dastidar has twenty years of model risk and AI governance experience at top-tier US banks, across all three lines of defense.

airisk-consulting.com
info@airisk-consulting.com
Virginia, USA

Provided for general information and does not constitute legal, regulatory or investment advice. AI Risk Consulting is not a law firm. The FSB sound practices report is a consultation and may change before it is finalized. Quotations from supervisory material are reproduced from the published documents; readers should consult the primary sources for authoritative text. © 2026 AI Risk Consulting.