AI Risk Architecture in Practice

A Practical Use Case Paper for Financial Services and Fintech Organizations

Date: May 2026 | Version 1.0

RIHO VILIPPUS
Senior Risk Manager & AI Governance Consultant

rihovilippus.com
consult@rihovilippus.com

1. The Problem This Paper Addresses

Most organizations approaching AI governance ask the wrong question. They ask, "***Do we have a governance framework?***" The right question is, "***Can we prove our governance is working today?***"

These are not the same question. One points to documentation. The other points to operating reality. The gap between the two is where actual risk lives - and in financial services, that gap is what regulators, auditors, and boards are now actively trying to measure.

This paper walks through what AI Risk Architecture looks like in practice - not as a policy exercise, but as an operational discipline. It is structured around a real scenario, uses concrete examples, and ends with actionable steps that a risk function can begin implementing without waiting for the next governance cycle.

2. What AI Risk Architecture Actually Is

AI Risk Architecture is the structural layer that connects AI use cases to controls, ownership, evidence, and escalation paths. It is not a framework document. It is the operating model that makes governance executable.

Think of it this way: a policy says ***what*** should happen. Architecture determines ***how*** it happens, ***who*** is accountable when it does not, and what evidence confirms it is working. Organizations that have built strong policy libraries, but weak architecture are, in practice, ungoverned - regardless of what the documentation says.

This distinction matters urgently in financial services. AI systems used for credit decisioning, fraud detection, customer-facing recommendations, and operational automation are subject to SR 11-7 model risk guidance, FINRA supervisory expectations, and the recently released Treasury FS AI RMF. Each of these frameworks requires more than written policies. They require demonstrable control effectiveness - controls that can be observed, tested, and traced to specific outcomes.

3. A Practical Scenario: The Mid-Size Fintech

Consider a mid-size fintech lending platform that deploys an AI credit scoring model. The risk team has:

- A model risk policy signed by the CRO
- A vendor due diligence checklist completed at onboarding
- A quarterly model review scheduled on the governance calendar

On paper, this looks controlled. In practice, the gaps are significant:

- **No continuous monitoring**: The model's performance is reviewed quarterly, but drift can occur within weeks of a distribution shift in the underlying data.
- **No ownership at runtime**: The policy names a model owner, but when the model's output begins to diverge, no one is operationally responsible for catching it between reviews.
- **No evidence trail**: The quarterly review produces a report, but there is no logging infrastructure that connects the report's conclusions to the model's actual daily outputs.

If a regulator examines this firm and asks, "***Can you demonstrate that your credit model performed within acceptable parameters last Tuesday?***" - the answer is no. Not because the team was negligent, but because the architecture was designed for documentation, not for operational governance.

4. Building the Architecture: Four Components

Effective AI risk architecture for a use case like the above requires four concrete components working together.

1. Use Case Classification and Risk Tiering

Not every AI system carries the same risk. A chatbot retrieving FAQ content is categorically different from a model influencing credit approvals - and the controls applied to each should reflect that gap. Before anything is designed, use cases need to be classified: at minimum by decision type (automated vs. advisory), impact scope (individual vs. systemic), and regulatory exposure.

The US Treasury's FS AI RMF provides a practical starting point, mapping AI adoption stages to applicable control objectives. Without tiering, organizations do one of two things: they apply uniform governance overhead to everything, which creates compliance fatigue, or they apply controls inconsistently and leave high-risk deployments under-resourced. Both outcomes produce real exposure.

2. Ownership That Is Live, Not Assigned Once

Governance without live ownership is decoration. Every AI use case in production needs a named owner who carries accountability between review cycles - not just at deployment, not just at the next audit, but continuously.

That means the owner receives monitoring signals, sits on escalation paths, and can authorize intervention. A model owner whose responsibility activates only when something goes wrong is not a governance control. They are a post-incident attribution mechanism. There is a significant difference between the two, and most governance programs are currently built around the second without acknowledging it.

3. Runtime Controls and Continuous Monitoring

This is the most structurally underfunded component in most organizations - and the one that matters most when a regulator shows up.

Runtime governance means controls that operate while the AI system is running. Not controls applied before launch and reviewed six months later. In the fintech lending scenario, this means automated performance threshold alerts covering accuracy, fairness metrics, and score distribution; output logging with retention aligned to regulatory requirements; and a defined escalation path that triggers when thresholds are breached - not when someone notices something looks off. FINRA's 2026 guidance treats prompt and output logging as a core supervisory requirement. Not a nice-to-have. Not a future consideration.

4. Evidence Architecture

Controls only have governance value if they leave evidence.

Evidence architecture means that the artifacts produced by monitoring, testing, and oversight are structured, retained, and traceable to specific dates and decisions. This is not about producing more reports. It is about designing the system so that when an examiner asks whether a control worked on a specific date, the answer can be retrieved - not reconstructed. Retrieval and reconstruction are not equivalent under examination. One confirms control effectiveness. The other confirms that someone worked hard to explain what probably happened.

5. Where Organizations Consistently Fail

Three failure patterns appear across financial services organizations regardless of size:

Failure Pattern	What It Looks Like	Actual Risk
Policy-architecture gap	Governance documents exist; no one can demonstrate the controls operate	Regulatory finding, reputational exposure
Static ownership	Model owner named at launch; no ongoing accountability mechanism	Undetected drift, unmanaged degradation
Review-cycle dependency	Governance activates at scheduled intervals only	Exposure between cycles, especially in fast-moving models

If your risk team cannot answer the question "Is our highest-risk AI system operating within acceptable parameters right now?" - the architecture work has not started yet.

The third pattern is particularly dangerous for AI systems. Traditional model risk management was designed for models that change slowly. AI systems - particularly those with continuous learning components or those exposed to distribution-shifting data - can degrade materially between quarterly reviews. A governance architecture built on review cycles alone is not calibrated to AI's operational tempo.

6. What Good Looks Like: A Practical Checklist

For a single AI use case in a regulated financial institution, the minimum viable architecture requires six artifacts - actively maintained, not filed and forgotten.

It starts with a **use case register entry** that captures risk tier, decision type, regulatory classification, and data lineage. That entry feeds directly into ownership: a **named owner** with a documented accountability scope and active enrollment in escalation paths. Not someone copied on a distribution list. Someone who receives a signal when something moves.

From there, the architecture needs a **performance monitoring dashboard** with defined thresholds, breach history, and a response log - and output and prompt logs retained in line with regulatory record-keeping requirements. These two work together. The dashboard tells you something changed. The logs tell you what happened.

- **Validation record** covering initial validation and any subsequent re-validation triggered by performance alerts or model changes
- **Evidence package** ready for regulatory examination - structured, retrievable, and not dependent on reconstruction

This is the minimum for one use case. Multiply it across an AI portfolio without systematic architecture and the governance burden becomes unmanageable. With architecture, it becomes repeatable.

7. The Core Argument

AI governance becomes real only when it stops being about what you have documented and starts being about what you can demonstrate. Documentation describes intent. Architecture produces evidence. Evidence is what regulators examine, what auditors test, and what boards should be asking for.

For financial services and fintech organizations deploying AI at scale, the question is not whether a governance framework exists. The question is whether it is operational - and whether, on any given day, someone in the organization can answer for every high-risk AI system running in production.

The organizations building that answer now will not be caught unprepared when the examiner asks.

About the Author

Riho Vilippus is a Senior Risk Manager and AI Governance and Risk Architecture Consultant. His work sits at the intersection of risk architecture, regulatory compliance, and AI deployment - translating abstract policy requirements into executable control structures that can be tested, monitored, and demonstrated to regulators. He brings hands-on experience across the full AI governance lifecycle: use case classification and risk tiering, model validation frameworks, vendor due diligence, and accountability design.

He advises organizations on closing the gap between governance documentation and governance that actually operates - live ownership models, runtime controls, continuous monitoring, and evidence architecture that survives examination. His advisory work spans US and EU regulatory environments, including SR 11-7 model risk guidance, the NIST AI RMF, the EU AI Act, and the US Treasury Financial Services AI Risk Management Framework.

His writing focuses on one problem: governance programs that hold up in the real world, not just on paper.

He is based in Europe and works with financial services and fintech organizations across the US and European markets.

Webpage: rihovilippus.com

Contact: consult@rihovilippus.com

LinkedIn: linkedin.com/in/rihovilippus