

“Governing AI Investment: A System-Level Assurance Framework for Financial Institutions”

Author: Ms. Ayasha Vadhyani [Research Scholar], Swami Ramanand Teerth Marathwada University, Nanded

Co-Author: Dr. R. B. Rampure [Guide], Peoples College, Nanded

Abstract - As the financial sector increasingly adopts artificial intelligence (AI), the development of governance frameworks has been lagging behind, leaving a very wide gap between the deployment of AI and the implication of its risks. The research is based on the 2025 EMEA Model Risk Management Survey. It covered survey of 87 banks and 49 insurance companies across Europe, the Middle East and South Africa in order to explore patterns of AI adoption, implementation challenges, and governance issues in financial institutions. Transparency and explainability continue to be top cited challenges, along with skills gaps and regulatory uncertainty. To fill these gaps, this paper suggests the framework of System-Level Intelligence Assurance, which is based on four pillars of architecture, shared memory, continuous validation loop, confidence governance, and auditable evidence chain. It is an extension to the concept of Model Risk Management that introduces agentic AI systems and replaces the old governance model of paperwork with one that is ongoing and built into the architecture. The results provide valuable insights for financial institutions, regulators, and risk practitioners aiming to safely scale AI while ensuring strong oversight and accountability.

Key Words: Artificial Intelligence Governance, Model Risk Management, Agentic AI, Financial Services, System-Level Assurance, and Regulatory Compliance.

1. INTRODUCTION

The financial services industry is rapidly encircling AI, yet governance is struggling to keep pace. The 2025 EMEA MRM Survey conducted by Deloitte which consisted of 87 banks and 49 insurers reveals that while 67% of banks and 57% of insurers now use AI/ML models, only 29% of banks and 15% of insurers consider their governance frameworks adequate. Traditional MRM frameworks which are built for static, interpretable models fail to address agentic AI systems that learn consistently and exhibit emergent behavior. This research proposes System-Level Intelligence Assurance. This is a framework grounded in four pillar shared memory, continuous validation loops, confidence governance, and auditable evidence chains. It extends traditional MRM to address the dynamics of agentic AI.

2. Literature review

This paper by Fritz-Morgenthal, Hein, and Papen Brock (2022) conducted a study on “Financial Risk Management and Explainable, Trustworthy, Responsible AI”. The study highlights at how banks and financial companies can safely use Artificial Intelligence (AI) along with following the rules. The

authors emphasized that the most important thing is not how complex a model is, but also what it is used for. For example, if an AI model is built to make decision for who gets a loan, it must be fair and easy to explain the reason behind such decision, no matter if it is a simple or a very advanced system. The article explains how regulatory bodies, including as the European Union, are now requiring AI systems to be open and transparent. The writers recommend adopting tools like SHAP to address the issue that some AI models are “black boxes”—that is, we are unable to see the way they make decisions. In order to make AI decisions explicit and intelligible, the study proposes utilizing SHAP (Shapley Additive explanations). It also emphasizes that in order to guarantee accountability, people must always have the last say.

The study by Djeflal, Siewert, and Wurster (2022) analyses AI strategies across 22 countries and the EU. The study reveals considerable variation in how states govern AI. Governance approaches span from self-regulation-promoting and market-based models to entrepreneurial and regulatory approaches. The study's focus on conclusion is that public responsibility for AI appears to be a cross-cutting issue that cannot be linked to any particular governance method, and it is essentially independent of the policy mix selected. The state whether embrace an active developmental role or a passive market-oriented approach, its commitment to responsible AI remains a separate and overarching problem.

3. Objectives of the Study

1. To analyse the use of AI by Financial Institutions.
2. To identify opportunities offered by AI-based investment technologies.
3. To examine challenges associated with AI.
4. To examine AI adoption, risks, and governance across EMEA banks and insurers.
5. To propose a novel System-Level Intelligence Assurance framework to bridge the gap between rapid AI deployment and inadequate risk controls.

4. Research Methodology:

This research is based on a survey conducted by Deloitte. The survey was carried out by Deloitte between July and September 2025, and included 136 organizations in Europe, the Middle East and South Africa, comprising 87 banks and 49 insurers. The survey utilized a detailed questionnaire that included questions on current practices in Model Risk Management (MRM), maturity ratings, primary obstacles, and future plans, with a particular emphasis on the adoption of AI and the potential risks. The sample consisted of large, medium and small institutions, ensuring a broad cross-section of the industry

landscape. The data gathered from the banking and insurance industry allowed for a comparison of the use of AI in fraud detection, customer experience and MRM automation, and helped identify governance gaps, regulatory hurdles and implementation challenges for the industry.

5. Use of AI by Deloitte

5.1. Model Risk Management Survey

The EMEA Model Risk Management (MRM) Survey was carried out in September 2025 with 87 banks and 49 insurers from across Europe, the Middle East and South Africa. The survey covered questions around their current practices in MRM, maturity, issues, and future endeavors by participants. It synthesizes the information into a core discussion about two main questions: how these institutions are integrating AI into their processes and what tangible challenges are preventing their adoption of legacy infrastructure, regulatory uncertainty, and governance issues.

The most common use cases for AI techniques in models are fraud detection (Anti-Money Laundering (AML) and Know-Your-Customer (KYC); 58% of banks and 30% of insurers use AI techniques in models) and Customer Experience (CX); 53% of banks and 37% of insurers use AI techniques in models).

The survey data shows that 60% of banks have enhanced their MRM processes for AI/ML, while just 29% find their processes are good enough. The disparity is even higher for insurers, with two-thirds having processes in place, and only 15% feeling they are effective. This gap in adoption represents an inability of traditional MRM concepts, designed for static, bounded and interpretable models, to stay in step with the systems they are supposed to manage.

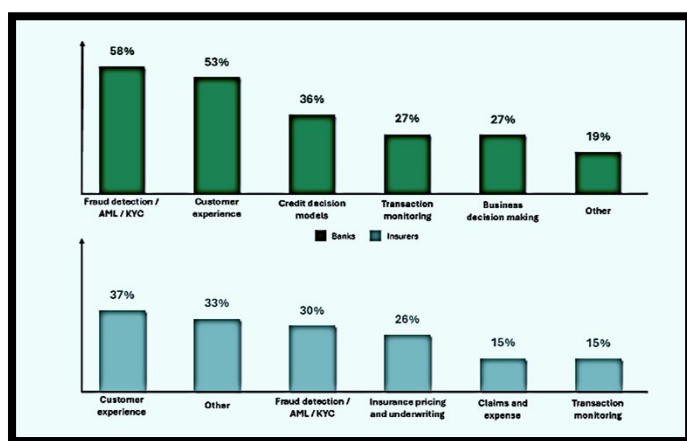


Fig -1: Top use cases for AI techniques (Deloitte's 2025 EMEA Model Risk Management Survey)

Given the regulatory mandate to perform Anti-Money Laundering (AML) and Know-Your-Customer (KYC) activities combined with the highly manual, repetitive, and rule-based nature of these processes the widespread adoption of AI in this space is hardly surprising. In practice, fraud detection models are used in two main ways: to get better at spotting real fraud and reducing false alarms, or to cast a wider net to catch new types of risk—such as those linked to cryptocurrency or environmental factors.

From the customer perspective, AI is popularly applied in predictive marketing, personalization and chatbots, which are perceived as 'low-barrier' channels. With the advent of GenAI, hyper-personalization has become even more possible as not

only the product selection is customized for an individual customer, but also the sales message itself.

Within the field of Model Risk Management, the discipline that governs how financial institutions validate and monitor their risk models 40% of participating banks are using or planning to use AI, compared to just 16% of insurers. Many AI use-cases, such as fraud detection, Customer Experience, or Document generation, are low-risk options that are picked as a starting point for AI adoption. But these applications have risks which are easy to be underestimated. They are not typically included in the "high-risk" category of the EU AI Act, but real-world examples since 2023 reveal otherwise: chatbots have breached confidentiality, provided misleading or illegal information, maliciously criticized their employers and even created backdoors for hackers. The questions asked and the answers received are very unpredictable; traditional up-front validation cannot detect the dangers of this.

5.2 Risk Implications of Common AI Applications

Many AI use-cases, such as fraud detection, Customer Experience, or Document generation, are low-risk options that are picked as a starting point for AI adoption. But these applications have risks which are easy to be underestimated. They are not typically included in the "high-risk" category of the EU AI Act, but real-world examples since 2023 reveal otherwise: chatbots have breached confidentiality, provided misleading or illegal information, maliciously criticized their employers and even created backdoors for hackers. The questions asked and the answers received are very unpredictable; traditional up-front validation cannot detect the dangers of this.

Over half of the respondents identified transparency and explainability as a challenge, particularly as vendors have increasingly introduced their solutions over in-house constructed AI asset and as its methodology has gotten more complex, from 'simple' ML tools to full Agentic AI workforces. The complexity also shows in the fact that 39% of the respondents said one challenge to AI implementation was the internal skills and capabilities. While willingness of employees to engage with AI was not cited as a significant hurdle, 24% of overall respondents cite rigidity of the existing processes as a blocker to using AI; however, only 3% of small banks. This indicates that larger institutions, with greater resources to dedicate to AI, do also have other challenges – as they must implement change in the current processes, and work on data quality and availability in current data systems. Further hurdles cited are fairness concerns (34% of respondents), safety and security (32%), regulatory landscape (34%) and generally risks posed by AI (46%) clearly showing, that risk management questions are far from answered, both from a technical and strategy perspective.

5.3 Governing AI Expansion: A Practical Roadmap for Banks and Insurers

It is not about widespread adoption of AI; it's about strategic use. When deciding on the appropriate use cases, consider three aspects: (1) organizational readiness: the level of readiness of the people and processes; (2) technical complexity: potential feasibility and requirements of the use case; and (3) risk control: whether risks are properly identified and managed. After building a use case, make sure it's deployed correctly by reviewing change management processes (including

stakeholder acceptance, user training, phased deployment and performance monitoring KPIs). Last, give local teams the freedom to create local AI solutions, and clarify an organizational wide governance framework throughout the lifecycle of AI. This agility and control balance are what make AI investments sustainable and deliver long-term returns.

6. Proposed Framework: System-Level Intelligence Assurance

The survey data shows that 60% of banks have enhanced their MRM processes for AI/ML, while just 29% find their processes are good enough. The disparity is even higher for insurers, with two-thirds having processes in place, and only 15% feeling they are effective. This gap in adoption represents an inability of traditional MRM concepts, designed for static, bounded and interpretable models. The failure mode has shifted. It is no longer about narrow errors in isolated models, but about systemic breakdowns: context bleeding across agents, reasoning that contradicts itself, decisions with no traceable lineage, and confidence eroding under fragmented data. In a world where trading agents talk to each other, fraud detectors share signals, and underwriting models learn collectively, governance cannot stop at the model boundary. Instead, it must encompass the entire multi-agent system. System-Level Intelligence Assurance lifts governance from process to architecture. Four pillars—shared memory, continuous validation, confidence governance, and auditable evidence chains which will result in closing the gaps the survey uncovered.

Pillar 1- Governance Through Shared Memory Architecture

The Problem: In existing multi-agent deployments, each agent has its own memory silo. A credit scoring agent is running for hours, without the knowledge or benefit of any fraud detection agent who has just gained a new insight on the risk pattern. Organizational intelligence is not stored anywhere. The survey highlights this disparity: 61% of banks say they have identified AI/ML models on their inventories, but only 35% have specific definitions for AI/ML models, indicating a lack of semantic infrastructure to regulate intelligence sharing across systems. **The Solution:** To enable system-level governance, we propose a governed shared memory architecture with a standardized interface for agents. It rests on a dual memory model: • Open-set memory – atomic facts from unstructured content, stored as vector embeddings with temporal anchoring, enabling retrieval-augmented prompts. • Schema-enforced memory – typed property values defined by organizational schemas, with confidence scores and provenance metadata, supporting querying and downstream integration. Organizational policies ensure that something can be stored, how it is typed and who has access to it. The result is that intelligence is not spread across agent silos but rather it is built up across the enterprise.

Pillar 2- Validation Loops—From Periodic to Continuous

The Problem. MRM has been using the periodic validation cycles – annual reviews, quarterly monitoring. In agentic systems, where the systems reprovision between validation cycles, material changes to behavior may happen before a formal redevelopment. This survey reflects this challenge: transparency/explainability is the greatest challenge for both banks and insurance companies (more than 50%). **The Solution.** System-level assurance takes periodic gates and incorporates continuous validation loops—runtime embedded

validation checks and runtime audit of reasoning against evidence. This can be implemented by means of:

- **Verification gates:** The AI verifies the accuracy of the data it has before proceeding to the next stage. Anything that appears fictitious or overly ambiguous is eliminated.
- **Automated guardrails:** Similar to safety barriers, these keep an eye on the AI while it is operating. The guardrails quickly detect any unexpected or dangerous behavior by the AI.
- **Ensemble checks:** Several instances of the same AI operate concurrently. Their outcomes are contrasted. The system marks it as requiring human review if they don't agree.

These mechanisms enable decisions to be based on easily available information and evidence, which is not provided by current MRM frameworks.

Pillar 3- Confidence Governance—Knowing What You Don't Know

The Problem. Agentic AI can have a fatal flaw: overconfidence. The authority of an agent with a proven track record is the same as that of a newly deployed agent -nothing about its actual performance is shown by its authority. Uncalibrated confidence makes fairness concerns (34%) and safety/security risks (32%) significant hurdles identified by the survey. **The Solution.** Confidence governance complements this by way of behavioral trust models that builds to demonstrated performance. Trust is built on demonstrated actions, it can be built without, and authority changes need human approval. There are 5 properties that need to be measured on a continuous basis:

Does the agent know what he doesn't know? — Calibration of uncertainty estimates.

Does it seek assistance at the proper times — appropriate escalation when the level of confidence is low?

Does it complete what it sets out to do? — Completion rates and task persistence.

Does it make decisions that it will have to make other decisions? — Decision stability.

Does it adjust to directions when given "no"? — Does it take direction?

This skeptical analysis of evidence evaluation is designed to assure confidence scores are realistic in terms of actual capability.

Pillar 4: Auditable Evidence Chains—From Fluency to Verifiability

The Problem: The most harmful is not incorrect answers, but strong answers with insufficient, absent, or distorted evidence. For AI-generated reports, the inquiry should be if there is clear and verifiable evidence behind each claim. Although document creation and compliance checking are among the most prevalent MRM applications (40% of banks), AI-generated reports can pose huge provenance risks.

The Solution. To assure at system level, a system must retain a clear and checkable trace of each AI decision. An auditable evidence chain is a trail that documents the raw data that an AI

has used, as well as the final decision it has arrived at. It's kept in a system that allows regulators and validators to look back at the process of the AI and confirm exactly how it made its decision. When the AI is confident, auditors can review the evidence to ensure it is accurate. This allows AI decisions to be transparent, accountable and verifiable.

Key approaches include trace-level diagnostics which capture not just what happened, but what data was attached, the context, and the reasoning behind each step.

7. Closing the Gap

The survey shows that institutions are adopting AI at a rate that outpaces governance development. The survey indicates that institutions are rushing to adopt AI at a rate they cannot keep up with establishing proper governance. This is not a reason to slow down AI adoption. It is a need to change the governing of AI. The shift between MRM to System-Level Intelligence Assurance isn't a rejection of existing frameworks but an extension to them, to cover the dynamics that agentic AI brings, which already exist in the production systems, both with or without governance. System-level assurance builds trust in AI systems while working. It moves governance from paperwork to continuous controls which are build-in. This is not an academic exercise, but more of necessity for banks and insurers that are under the controlled environment where accountability is non-negotiable. The issue is not if such capabilities can be developed, but how fast they can be put into practice.

8. Conclusion

The 2025 EMEA MRM Survey reveals a financial services industry is rapidly embracing AI. But the industry is struggling to govern it effectively. The adoption rates have surge with 67% of banks and 57% of insurers now using AI/ML models, however a significant governance gap persists. Only 29% of banks and 15% of insurers consider their MRM frameworks adequate for AI oversight. The main hurdles include transparency challenges, skills shortages, and regulatory uncertainty.

To eliminate this gap, a proposal of System-Level Intelligence Assurance—a paradigm shift from model-level to system-level governance is been suggested. This is occupied in four pillars:

- Shared Memory – Instead of fragmenting in centralized structures, it allows intelligence to build up across agents.
- Continuous Validation Loops – substitutes runtime oversight for recurring reviews.
- Confidence Governance – Ensures trust is earned through displayed behavior, not assumptions.
- Auditable Evidence Chains – Makes every decision verifiable, not just fluent.

Banks, insurers, and regulators are guided to not use governance challenges as a reason to slow the adoption of AI. Instead, they should treat the challenges as a reason to upgrade in the way they govern AI. System-level assurance builds trust directly into the technology itself which will make it continuous and measurable.

REFERENCES

1. Fritz-Morgenthal S, Hein B and Papen Brock J (2022) Financial Risk Management and Explainable, Trustworthy, Responsible AI. *Front. Artif. Intell.* 5:779799. Doi: 10.3389/frai.2022.779799
2. Deloitte. (2025). EMEA Model Risk Management Survey 2025 (Banks and Insurers report).
3. Kurshan, E., Balch, T., & Byrd, D. (2025). The Agentic Regulator: Risks for AI in Finance and a Proposed Agent-based Framework for Governance. *arXiv*.
4. Deloitte India. (2026). Building Trust in Agentic AI: TRACE Framework for Policy-Driven Multi-Agent System Design. *Zenodo*.
5. Djeflal, C., Siewert, M. B., & Wurster, S. (2022). Role of the state and responsibility in governing artificial intelligence: a comparative analysis of AI strategies. *Journal of European Public Policy*, 29(11), 1799–1821. <https://doi.org/10.1080/13501763.2022.2094987>