*Just in Time*

# Model Risk Management of GenAI Workflows

February 2026

iason ESSENTIAL SERVICES FOR FINANCIAL INSTITUTIONS

# Executive Summary

The paper[1] examines how **generative artificial intelligence** (**GenAI**) fundamentally expands the **model risk surface** in financial institutions and challenges established **model risk management** (**MRM**) frameworks. While GenAI offers significant efficiency and productivity gains, it introduces **novel failure modes**, including open-ended outputs, composable architectures, reliance on heuristic design choices, vendor opacity, and continuous change without formal redevelopment.

[1] *ATI & Others, "Move Fast Without Breaking the Bank: Model Risk Management of GenAI workflows", January 2026*

# At a Glance

**Keywords:** Model Risk, Model Risk Management, GenAI, LLMs

01

# Introduction

Overview

# Introduction
## Overview

The GenAI is to be considered within the long tradition of quantitative modeling in financial institutions, where models have historically supported risk management, pricing, stress testing, and financial crime prevention. GenAI represents a qualitative shift: models are no longer confined to narrow numerical tasks but are increasingly deployed to **augment or replace human judgment** across a wide range of banking activities.

This expansion introduces significant **strategic and operational risks**. When novel modeling techniques are applied to unfamiliar tasks, the potential for **unrecognized or poorly understood risks** increases materially. Model risk management therefore becomes a central control function, ensuring that decisions based on model outputs remain reliable, explainable, and aligned with business objectives.

The paper reviews existing regulatory definitions of model risk and models, emphasizing the enduring relevance of U.S. Federal Reserve's SR 11-7's three-component model framework (inputs, processing, reporting). However, it highlights that **modern GenAI systems strain these definitions**, particularly where:

- Outputs are unstructured and probabilistic,
- Boundaries between data risk and model risk are blurred,
- Behavior is materially shaped by non-model artefacts such as prompts, retrieval indices, or orchestration logic.

The paper illustrates these challenges through common deployment patterns:

- **Retrieval-Augmented Generation** (**RAG**), where outputs depend on external knowledge bases and retrieval logic.
- **Fine-tuned and domain-specialized LLMs**, where even minor adaptation can significantly alter behavior.
- **Agentic workflows**, where models autonomously plan and execute sequences of actions, introducing emergent behaviors and compounding errors.

These patterns underscore that GenAI systems are rarely standalone; they are **composable, dynamic and vendor-dependent**. As a result, traditional static validation and change-management processes are insufficient. While GenAI implicates a wide range of risk types (operational, legal, ICT, conduct), the paper focuses specifically on **model risk management**, while recognizing the need for close coordination across control functions.

02

# Model Risk Classification and Tiering

General Model Governance

Model Identification and Risk Tiering & Change Management

# Model Risk Classification and Tiering 1/2

## General Model Governance

This section reiterates the foundational governance principles embedded in bank MRM policies, aligned with U.S. Federal Reserve's SR 11-7 and the UK PRA's SS1/23. These principles define clear roles across the **three lines of defense**, emphasizing accountability, independence and proportionality.

### First line of defense (1LoD)

The **first line of defense** (**1LoD**) owns GenAI use cases and is responsible for end-to-end model lifecycle management. For GenAI, this responsibility extends beyond traditional model development to include:

- Selection of architectures and vendors,
- Definition of intended use and outputs,
- Documentation of prompts, retrieval logic, tools, and dependencies,
- Design and execution of testing and monitoring frameworks,
- Establishment of fallback and contingency mechanisms.

Given the heuristic and vendor-driven nature of GenAI, developers must also maintain **dependency registers** capturing external APIs, models and knowledge bases.

### Second line of defense (2LoD)

The **second line of defense** (**2LoD**) provides independent challenge through validation. Validators assess conceptual soundness, empirical performance, robustness, documentation quality and monitoring adequacy. Independence is preserved by avoiding de facto co-development. Validation outcomes are formalized in **model validation reports**, which document findings, limitations and approval conditions.

The section emphasizes that GenAI does not alter the fundamental governance structure but **raises expectations** around documentation, evidence and testing rigor.

# Model Risk Classification and Tiering 2/2

## Model Identification and Risk Tiering & Change Management

### Model Identification and Risk Tiering

Identifying GenAI models presents a significant operational challenge due to their **pervasive and sometimes informal use** across business areas not traditionally accustomed to MRM. The paper recommends treating the **entire GenAI workflow** as the primary governed object, with subcomponents explicitly recorded.

Traditional tiering approaches - decision trees and scorecards - remain relevant but must be **augmented with GenAI-specific risk drivers**, including:

- Degree of autonomy,
- Sensitivity of content and data,
- Volatility of non-model artefacts,
- Evaluation uncertainty and robustness.

High-autonomy agents and external-facing systems with volatile knowledge bases should default to higher risk tiers, triggering deeper validation and monitoring. Tiering outcomes should directly determine the **depth of validation, testing requirements and monitoring cadence**, ensuring proportionality and consistency.

### Change Management

Change management is particularly challenging for GenAI due to **continuous evolution without formal redevelopment**. Updates to prompts, retrieval corpora, APIs, or guardrails can materially alter behavior and must be treated as **model changes**.

The paper stresses the importance of:

- Advance notification of material changes,
- Retention of artefacts for impact assessment,
- Integration of vendor updates into model inventories and validation cycles.

Vendor management is highlighted as a critical extension of existing third-party risk practices, with added emphasis on transparency, versioning, and software supply-chain security. Institutions are encouraged to contractually require documentation akin to **model cards and datasheets**, covering intended use, limitations and evaluation evidence.

# 03

## Model Design

Establishing Fitness Metrics & Loss Functions and Risk Metrics

Sensitivity Analysis, Assessment of Data Quality and Relevance
& Uncertainty Quantification

Just in Time

# Model Design 1/2

## Establishing Fitness Metrics & Loss Functions and Risk Metrics

Model design is framed as a core pillar of conceptual soundness under supervisory guidance. The paper emphasizes that while GenAI relies on advanced mathematics, **many critical design decisions are heuristic**, lacking strong theoretical justification. This increases the burden on documentation, testing, and independent challenge.

Developers must explicitly justify:

- Prompt templates and orchestration logic,
- Alignment strategies such as Reinforcement Learning from Human Feedback (RLHF),
- Data filtering and tuning decisions,
- Use of vendor-provided components.

Robust documentation is essential to enable independent replication and challenge. The section identifies five areas where standard MRM practices require uplift.

### Establishing Fitness Metrics

Fitness metrics define what "good performance" means for a GenAI system. Misaligned metrics create foundational model risk. The paper distinguishes between:

- **Reference-based metrics** (e.g., overlap and semantic similarity),
- **Reference-free metrics** (e.g., fluency, toxicity),
- **LLM-as-judge approaches**, which scale qualitative assessment but require calibration and uncertainty reporting.

For financial use cases, metrics must reflect **domain-specific risks**, such as factual accuracy, compliance and customer protection. Static benchmarks are insufficient; dynamic, task-specific evaluations are required to capture evolving risks.

### Loss Functions and Risk Metrics

The paper highlights a critical disconnect between **training losses** and **risk-relevant performance metrics**. While LLMs are trained to optimize generic objectives, governance requires evaluation against application-specific risk measures.

A defensible evaluation stack combines:

- Reference-based factuality where possible,
- Reference-free production monitors,
- Calibrated LLM-based evaluators with uncertainty,
- Reasoning and tool-use checks at the system level.

# Model Design 2/2

## Sensitivity Analysis, Assessment of Data Quality and Relevance & Uncertainty Quantification

### Sensitivity Analysis

Sensitivity analysis is essential for establishing operating boundaries. For GenAI, this extends beyond input perturbations to include:

- Prompt variation,
- Retrieval noise,
- Tool availability and schema drift,
- Reasoning path instability.

The paper recommends documenting a clear **operational envelope**, identifying validated regimes and conditions requiring guardrails or human oversight.

### Assessment of Data Quality and Relevance

Data quality is central to model reliability. The paper disaggregates the GenAI data lifecycle, noting that while pretraining data is often opaque, fine-tuning and alignment data are directly governed and must be rigorously assessed.

Key risks include:

- Distribution shift and recency decay,
- Spurious correlations,
- Embedded bias,
- Adversarial poisoning and backdoors.

Controls should include dataset documentation, provenance tracking, bias audits, and versioned snapshots linked to model checkpoints.

### Uncertainty Quantification

Uncertainty quantification (UQ) is identified as one of the most critical and challenging aspects of GenAI MRM. The paper surveys a range of approaches, from token-level confidence to ensemble methods and emerging conformal prediction techniques.

From a supervisory perspective, the objective is not to eliminate uncertainty but to **demonstrate that it is understood, measured and controlled**, with clear escalation and fallback mechanisms.

04

# Model Testing and Monitoring

Assessing Quality and Relevance of Test Data, Understanding Coverage of Edge-Case Risk with Stress Testing & Model Monitoring

# Model Testing and Monitoring

## Assessing Quality and Relevance of Test Data, Understanding Coverage of Edge-Case Risk with Stress Testing & Model Monitoring

### Assessing Quality and Relevance of Test Data

Given the empirical nature of GenAI evaluation, the choice of test data is paramount. Benchmarks provide useful baselines but must be carefully aligned with business use cases. The paper categorizes benchmarks into:

- Domain-specific,
- General capability,
- Safety-focused.

Institutions should combine these sources to demonstrate both task alignment and broader robustness.

### Understanding Coverage of Edge-Case Risk with Stress Testing

Static benchmarks are insufficient for capturing rare but material failures. The paper advocates **adversarial stress testing**, ranging from statistical perturbations to formal certification approaches.

Stress testing should be explicitly tied to business-relevant risks and include robustness, explainability, privacy and fairness dimensions where material.

### Model Monitoring

Monitoring is framed as the bridge between development and live operation. Effective monitoring requires:

- Business-aligned risk metrics,
- Thresholds and escalation paths,
- Continuous assessment of drift, uncertainty and dependency changes.

Future-proofing performance requires ongoing re-evaluation as data, tools and external conditions evolve.

# 05

## Concluding Remarks

# Concluding Remarks

Just in Time

a. **GenAI materially expands the model risk surface** through composability, dynamism and open-ended outputs.

b. **Existing supervisory frameworks (U.S. Federal Reserve's SR 11-7 and the UK PRA's SS1/23) remain valid**, but require enhanced implementation to address GenAI-specific risks.

c. **Whole-workflow governance** is essential; non-model artefacts must be treated as first-class risk drivers.

d. **Risk tiering must incorporate GenAI-specific factors**, including autonomy, dependency volatility and evaluation uncertainty.

e. **Heuristic design choices require explicit justification, testing and documentation** to meet conceptual soundness standards.

f. **Static benchmarks are insufficient**; dynamic, task-specific evaluation and adversarial stress testing are necessary.

g. **Uncertainty quantification is a critical control**, enabling calibrated use, escalation and fallback mechanisms.

h. **Continuous monitoring and disciplined change management** are essential to maintain alignment between model behavior, business intent and supervisory expectations.

# ESSENTIAL SERVICES FOR FINANCIAL INSTITUTIONS

**iason** is an international consulting firm that has been supporting both financial institutions and regulators in topics related to Risk Management, Finance and ICT since 2008

## Strategy

**Strategic advisory** on the **design** of **advanced frameworks** and **solutions** to fulfil both **business** and **regulatory needs** in Risk Management and IT departments

## Methodology & Governance

**Implementation** of the designed **solutions** in bank departments **Methodological support** to both **systemically important financial institutions** and **supervisory entities**

## Solution

Advanced **software solutions** for **modelling, forecasting, calculating** metrics and **integrating** risks, all on cloud and distributed in Software-as-a-Service (**SaaS**)

KEEP IN TOUCH

# Company Profile

**iason** is an international firm that consults Financial Institutions on Risk Management.
**iason** integrates deep industry knowledge with specialised expertise in Market, Liquidity, Funding, Credit and Counterparty Risk, in Organisational Set-Up and in Strategic Planning.

**Dario Esposito**

**Marco Musto**

**Tommaso Travenzoli**

www.iasonltd.com