International Journal of Artificial Intelligence, Data Science, and Machine Learning
Grace Horizon Publication | Volume 7, Issue 1, 56-70, 2026
ISSN: 3050-9262 | https://doi.org/10.63282/3050-9262.IJAIDSML-V7I1P113

Original Article

Third-Party Model Risk: Advanced Due Diligence and Contractual Oversight for Embedded AI/ML Solutions in SaaS Core Banking and Risk-as-a-Service Platforms -- A Model Governance and Regulatory Risk Framework for Financial Institutions.

Puneet Redu
Independent Researcher, USA.

Received On: 14/12/2025 ***Revised On: 16/01/2026*** ***Accepted On: 23/01/2026*** ***Published On: 30/01/2026***

Abstract *- The transformation of financial institutions toward modular, platform-based operating models has altered how quantitative and algorithmic decision systems are developed, deployed, and governed. Software-as-a-Service (SaaS) core banking platforms and Risk-as-a-Service (RaaS) providers increasingly embed externally developed Artificial Intelligence (AI), Machine Learning (ML), and quantitative models into functions such as credit underwriting, fraud detection, capital estimation, and regulatory reporting. While this shift offers meaningful efficiency and analytical benefits, it also introduces a structurally distinct form of model risk driven by external control, limited transparency, continuous vendor-managed change, and increasing concentration on a small number of technology providers.Existing regulatory frameworks including the Federal Reserve's SR 11-7, the UK Prudential Regulation Authority's SS1/23, the EU's Digital Operational Resilience Act (DORA), and the Monetary Authority of Singapore's Technology Risk Management (TRM) Guidelines establish that institutions retain responsibility for the governance and risk management of third-party models [1–4]. However, these frameworks are intentionally principle-based and provide limited operational guidance on how to govern, validate, and evidence effective challenge over opaque, externally operated models in practice.*

This paper addresses that gap by developing a unified governance and validation framework specifically designed for embedded third-party AI/ML and quantitative models. It makes three primary contributions. First, it formalizes third-party model risk as a distinct category of model risk characterized by structural features that differ materially from those of internally developed models, particularly with respect to transparency, control, and concentration [5,6]. Second, it proposes a lifecycle-based governance and black-box validation framework that enables independent challenge, performance monitoring, and regulatory defensibility even where access to source code, training data, or internal model logic is limited [7]. Third, it integrates legal, audit, and technical controls into a single operational approach, translating high-level supervisory expectations into enforceable contractual rights, audit evidence standards, and validation practices [3,4].By shifting institutions from passive reliance on vendor assurances toward active, evidence-based governance of embedded analytical systems, the framework supports regulatory compliance, operational resilience, and systemic stability in an increasingly platform-driven financial ecosystem [5,6].

Keywords *- Model Risk Management (MRM), Artificial Intelligence, Machine Learning, Saas Core Banking, Risk-As-A-Service (Raas), SR 11-7, PRA SS1/23, DORA, MAS TRM, Validation under Opacity.*

1. Introduction

The increasing reliance of financial institutions on quantitative models and algorithmic decision systems has reshaped the structure of modern banking. Models are no longer confined to specialist risk functions or analytical support teams; they now operate at the core of business processes, influencing credit approvals, transaction monitoring, capital calculations, liquidity management, and regulatory reporting. Over the past decade, this reliance has been reinforced by a shift in technology architecture from vertically integrated, institution-controlled systems toward modular, platform-based operating models. Core banking functions, risk analytics, and compliance capabilities are increasingly delivered through Software-as-a-Service (SaaS) platforms and managed service providers, often referred to collectively as Risk-as-a-Service (RaaS).

This architectural shift has important implications for how model risk is created, transmitted, and governed. In traditional environments, institutions typically developed or

directly implemented their own models, retained access to source code and documentation, and exercised direct control over model change, deployment, and monitoring. In contrast, under platform-based models, institutions embed externally developed Artificial Intelligence (AI), Machine Learning (ML), and quantitative systems into operational workflows while relying on vendors to maintain, update, and sometimes even retrain those systems. Responsibility for outcomes, however, remains firmly with the institution.

Regulatory frameworks across major jurisdictions have responded by clarifying that accountability for model risk cannot be outsourced. Supervisory guidance such as the Federal Reserve's SR 11-7, the UK Prudential Regulation Authority's SS1/23, the EU's Digital Operational Resilience Act (DORA), and the Monetary Authority of Singapore's Technology Risk Management (TRM) Guidelines all affirm that institutions remain responsible for the design, performance, and risk impacts of models, including those developed or operated by third parties [1–4]. These frameworks establish core principles of governance, validation, and oversight, and they explicitly bring third-party models within scope.

At the same time, these regulatory standards are intentionally principle-based. They describe *what* institutions are accountable for but leave considerable discretion in *how* accountability is achieved in practice. This flexibility is appropriate given the diversity of institutions and technologies, but it also creates practical challenges in environments where models are externally controlled, algorithmically opaque, and continuously evolving. Traditional model risk management approaches assume a level of transparency, documentation, and operational control that may be difficult to obtain when models are proprietary, embedded within vendor platforms, and updated outside the institution's direct control.

As a result, institutions face a structural tension between responsibility and control. They are expected to evidence effective challenge, ongoing validation, and governance over systems that they do not fully design, cannot fully inspect, and cannot always directly test. This tension is not merely operational; it has regulatory, legal, and systemic dimensions. It affects the institution's ability to demonstrate compliance, to respond to supervisory scrutiny, to manage operational resilience, and to understand how correlated dependencies on common vendors may propagate risk across the financial system.

This paper addresses that tension by developing a governance and validation framework specifically tailored to third-party AI/ML and quantitative models embedded within SaaS and RaaS platforms. Rather than treating third-party model risk as a simple extension of either vendor risk or internal model risk, the paper treats it as a distinct configuration of risk characterized by external control, limited transparency, dynamic change, and concentration. These features alter both the nature of model risk and the tools required to manage it.

The analysis proceeds in three steps. First, it examines the evolving regulatory landscape to clarify the expectations placed on institutions with respect to externally developed models and third-party platforms. Second, it develops a technical and governance-oriented risk taxonomy that highlights how embedded AI/ML systems introduce specific vulnerabilities related to explainability, data provenance, performance stability, and systemic concentration. Third, it proposes a unified framework that integrates governance structures, validation methodologies, audit practices, and contractual controls into a single operational approach to third-party model risk.

The central argument of the paper is that effective governance of embedded third-party models requires moving beyond reliance on vendor assurances and high-level policy statements toward evidence-based, operationally enforceable controls. This includes not only technical validation and monitoring, but also the design of contractual rights, audit mechanisms, and escalation processes that enable institutions to demonstrate accountability in environments where direct transparency is limited. By articulating and operationalizing this approach, the paper aims to contribute to a more robust and resilient model risk discipline that reflects the realities of platform-based financial infrastructure.

2. The Global Regulatory Landscape for Third-Party Model Risk

The governance of quantitative and algorithmic models in financial institutions has historically been shaped by a set of supervisory frameworks that emphasize accountability, soundness, and independent oversight. Although these frameworks were initially developed in response to risks arising from internally developed models, they have evolved to encompass models developed, operated, or embedded by third parties. Across jurisdictions, regulators have converged on the principle that institutions remain fully responsible for the risks created by their use of models, regardless of whether those models are built internally or sourced externally.

This section examines four influential regulatory frameworks — SR 11-7 in the United States, SS1/23 in the United Kingdom, DORA in the European Union, and the MAS Technology Risk Management and outsourcing guidelines in Singapore to highlight both the common foundations and the areas where operational guidance remains underdeveloped for third-party model governance.

2.1. Federal Reserve SR 11-7: Accountability and Effective Challenge

The Federal Reserve's Supervisory Guidance on Model Risk Management (SR 11-7) defines a model broadly as a quantitative method that applies statistical, economic, financial, or mathematical theories to process input data into quantitative estimates. Importantly, SR 11-7 explicitly states that a bank's responsibility for model risk does not depend on whether a model is developed internally or obtained from a vendor. Institutions are expected to understand the models

they use, assess their limitations, and manage their risks accordingly [1].

A central concept in SR 11-7 is "effective challenge," defined as the critical analysis of a model's design, assumptions, implementation, and outputs by informed and objective parties. Effective challenge is intended to prevent unquestioned reliance on model outputs and to surface weaknesses before they lead to adverse outcomes. While this concept is well-developed for internally built models, its application becomes more complex when models are externally developed and proprietary. In such cases, institutions may not have access to source code, training data, or internal design documentation, yet they remain expected to demonstrate understanding and control.

SR 11-7 does not prescribe specific methods for achieving effective challenge in these circumstances. Instead, it leaves institutions to determine how to obtain sufficient assurance over vendor models through a combination of due diligence, validation, monitoring, and governance. This flexibility allows adaptation to different technologies, but it also creates variability in practice and uncertainty about what constitutes sufficient evidence of effective challenge for externally controlled systems.

2.2. PRA SS1/23: Model Risk as a Distinct Risk Discipline

The UK Prudential Regulation Authority's SS1/23 builds on earlier supervisory statements by explicitly framing model risk management as a risk discipline in its own right [2]. It sets out five core principles covering model identification and classification, governance and oversight, development and implementation, independent validation, and model risk mitigants. SS1/23 applies to all models that materially influence decision-making, including those developed or operated by third parties.

SS1/23 reinforces the idea that institutions must take a strategic and holistic approach to model risk, embedding it within enterprise risk management rather than treating it as a purely technical function. It also emphasizes the need for proportionality: more complex and higher-impact models require more intensive governance, validation, and oversight.

For third-party models, SS1/23 implies that institutions should classify vendor models according to materiality and risk, ensure that appropriate governance structures apply, and maintain the ability to challenge and monitor those models over time. As with SR 11-7, however, SS1/23 remains principle-based. It articulates expectations but does not provide detailed operational guidance on how to validate opaque models, how to manage continuous vendor-driven change, or how to structure contractual arrangements to support governance objectives.

2.3. DORA: Third-Party Risk and Operational Resilience

The EU's Digital Operational Resilience Act (DORA) introduces a more prescriptive framework for managing risks arising from information and communication technology (ICT), including risks related to third-party service providers [3]. DORA reflects a growing regulatory concern that concentration on a small number of critical service providers could create systemic vulnerabilities.

Under DORA, institutions are required to identify and manage ICT-related risks, conduct due diligence on third-party providers, and ensure that contractual arrangements support resilience, auditability, and supervisory access. Contracts are expected to include provisions relating to performance monitoring, incident reporting, audit and inspection rights, and exit strategies.

While DORA does not focus specifically on model risk, its contractual and resilience requirements have direct implications for third-party models embedded within ICT platforms. In effect, DORA elevates third-party governance from a bilateral commercial matter to a prudential concern, reinforcing the need for institutions to design contracts that support not only service continuity but also risk governance and supervisory oversight.

2.4. MAS TRM: Technology Risk and Outsourcing Governance

The Monetary Authority of Singapore's Technology Risk Management guidelines and outsourcing requirements similarly emphasize due diligence, ongoing monitoring, and contractual controls over third-party service providers [4]. MAS expects institutions to assess the risks posed by outsourcing arrangements, ensure that service providers meet security and resilience standards, and retain sufficient control to manage risks effectively.

For embedded AI/ML and quantitative models, this implies that institutions should understand how vendor systems operate, how data are used and protected, and how changes are managed over time. As with the other frameworks, MAS emphasizes accountability and resilience but leaves institutions to design the specific mechanisms through which these objectives are achieved.

2.5. Convergence and Remaining Gaps

Taken together, these frameworks reflect a strong convergence on three principles. First, institutions remain accountable for the risks created by models, regardless of whether those models are internal or external. Second, governance, validation, and oversight are essential components of responsible model use. Third, third-party dependencies are increasingly recognized as sources of operational and systemic risk.

At the same time, a common feature across these frameworks is their high-level nature. They establish expectations but do not specify how institutions should validate black-box models, monitor continuously evolving vendor systems, evidence effective challenge without full transparency, or align legal, technical, and audit controls into a coherent governance approach. These unresolved questions are not deficiencies of the frameworks; rather, they reflect the pace of technological change and the diversity of institutional contexts.

The purpose of this paper is not to critique these frameworks but to extend them by translating their principles into operational mechanisms suitable for platform-based financial infrastructures. The framework proposed in the following sections is intended to complement, not replace, existing regulatory guidance by providing institutions with practical tools to demonstrate accountability, resilience, and effective challenge in environments characterized by external control and limited transparency.

3. Risk Taxonomy of Embedded Third-Party Models

The increasing use of externally developed and operated AI/ML and quantitative models within financial institutions introduces a set of risks that differ in important ways from those associated with internally developed models. These differences do not arise primarily from the mathematical form of the models themselves, but from their structural context: external ownership, proprietary design, continuous vendor-managed change, and embeddedness within operational platforms. This section develops a taxonomy of third-party model risk that highlights these structural and technical dimensions and explains how they interact to create distinctive risk profiles. Academic and supervisory work has highlighted how opacity, external control, and concentration change the nature of model risk relative to traditional internal settings [5,6]

3.1. Structural Risk Drivers

3.1.1. External Control and Limited Transparency

Externally developed models are typically treated by vendors as proprietary intellectual property. As a result, institutions may have limited access to source code, training data, model architecture, or design documentation. This constrains traditional validation practices, which rely on detailed inspection of model logic and assumptions. Instead, institutions must infer model behavior from inputs and outputs, increasing reliance on indirect forms of evidence.

Limited transparency also affects governance and accountability. When model changes are implemented by vendors, institutions may receive only high-level descriptions of updates, making it difficult to assess the impact of changes on model performance, fairness, stability, or regulatory compliance. This creates a dependency on vendor disclosures and processes that may not align fully with supervisory expectations.

3.1.2. Continuous Vendor-Managed Change

Unlike many internally developed models, which are updated episodically and under institution-controlled change management processes, vendor models are often updated continuously as part of product development cycles. These updates may reflect improvements, bug fixes, data refreshes, or algorithmic changes, but they can also introduce new risks.

Continuous change complicates validation and monitoring. A model that has been validated at one point in time may evolve in ways that render prior validation partially obsolete. This dynamic undermines the traditional assumption that validation is a periodic activity and instead requires ongoing surveillance of model behavior.

3.1.3 Concentration and Correlated Dependency

Platform-based delivery models create incentives for standardization and scale, leading many institutions to rely on a small number of dominant vendors. This concentration can generate correlated risk: if multiple institutions use similar models or platforms, weaknesses or failures in those systems can propagate across the financial system.

This risk is not limited to operational outages. It can also arise through synchronized decision-making, where models trained on similar data and optimized for similar objectives generate correlated responses to market conditions, amplifying volatility or reinforcing systemic trends.

3.2. Technical Risk Dimensions

3.2.1. Explainability and Interpretability

Advanced AI/ML models, particularly deep learning and ensemble methods, often exhibit high predictive performance at the cost of interpretability. In regulated contexts, this trade-off creates challenges for accountability, fairness, and regulatory justification. Institutions must be able to explain how and why models produce particular outcomes, especially when those outcomes affect customers, capital, or compliance.

For externally developed models, explainability is further constrained by limited access to model internals. Institutions must rely on post-hoc explanation techniques or vendor-provided summaries, which may not fully capture model behavior or limitations.

3.2.2. Data Provenance and Representativeness

Model performance and integrity depend critically on the quality and relevance of training and input data. In third-party settings, institutions may have limited visibility into how training data were sourced, processed, and curated. This raises questions about representativeness, bias, legal compliance, and ongoing relevance as market conditions change.

If training data do not reflect the institution's specific portfolio, customer base, or operating environment, model outputs may be systematically biased or misaligned with risk appetite. Detecting such misalignment requires careful outcomes analysis and benchmarking rather than reliance on model design documentation.

3.2.3. Performance Stability and Drift

Models may degrade over time as data distributions shift, behaviors change, or external conditions evolve. For vendor models, institutions may not control retraining schedules, feature updates, or data refresh processes. As a result, performance drift may occur without clear visibility into its causes.

Detecting drift under these conditions requires continuous monitoring of outputs, stability metrics, and business impacts. It also requires defining what constitutes acceptable variation versus material degradation a judgment that depends on context, risk appetite, and regulatory expectations.

3.2.4. Security, Privacy, and Data Leakage

Embedded models often process sensitive financial and personal data. In SaaS environments, this data may traverse shared infrastructure or be processed alongside data from other clients. This raises risks related to confidentiality, data leakage, unauthorized use, and regulatory compliance with data protection laws.

Institutions must therefore consider not only model performance but also how data are handled within vendor systems, how access is controlled, and how incidents are detected and managed.

3.3. Interaction of Structural and Technical Risks

The most significant risks arise not from any single dimension, but from the interaction of structural and technical factors. For example, limited transparency combined with continuous change can undermine the effectiveness of periodic validation. Concentration combined with synchronized decision-making can create systemic vulnerabilities. Data opacity combined with regulatory accountability can expose institutions to legal and reputational risk.

This interactional nature of risk reinforces the need for an integrated governance approach. Addressing technical risks in isolation is insufficient if contractual arrangements, governance structures, and audit mechanisms do not support ongoing oversight and effective challenge. Conversely, strong governance without technical monitoring may fail to detect subtle but material shifts in model behavior.

By articulating these risk dimensions and their interactions, this taxonomy provides a foundation for the governance and validation framework developed in the subsequent sections. It clarifies why third-party model risk cannot be treated simply as a subset of vendor risk or internal model risk, and why it requires tailored tools, processes, and controls.

4. Governance Framework for Third-Party Models

The governance of third-party AI/ML and quantitative models embedded within financial platforms requires a shift from traditional, internally focused model risk management toward an approach that is explicitly designed for environments characterized by external control, limited transparency, and continuous change. This section proposes a governance framework that translates high-level regulatory expectations into operational structures and processes that enable institutions to exercise accountability, oversight, and effective challenge over externally developed and operated models.

The framework is built around three principles. First, governance must follow impact: models that materially influence decisions or risk profiles require governance commensurate with their potential consequences, regardless of where they are developed. Second, governance must be continuous rather than episodic, reflecting the dynamic nature of vendor-managed systems. Third, governance must be integrative, linking technical validation, organizational oversight, auditability, and contractual enforceability into a coherent whole.

The framework is operationalized through a lifecycle structure that applies governance controls from the point of vendor selection through model onboarding, deployment, operation, and eventual exit. This lifecycle perspective integrates preventative controls, behavioral validation, continuous monitoring, independent oversight, and contractual enforceability into a single control loop that supports accountability under conditions of external control and limited transparency. The structure of this lifecycle governance framework is summarized in Figure 1.

4.1. Lifecycle-Based Governance

The proposed framework adopts a lifecycle perspective in which governance is applied from initial selection through deployment, operation, and eventual exit.

4.1.1. Model Identification and Classification.

Institutions should explicitly identify all externally developed models that materially influence decisions or risk outcomes and include them within the model inventory. These models should be classified by materiality, complexity, and potential impact, with higher-risk models subject to more intensive governance and oversight.

4.1.2. Pre-Engagement Due Diligence

Before onboarding a vendor model, institutions should conduct structured due diligence covering conceptual alignment, data relevance, operational integration, security, and regulatory implications. This includes assessing whether the model's design assumptions are compatible with the institution's products, customers, and risk appetite, and whether the vendor's development and testing practices meet acceptable standards.

4.1.3. Governance and Ownership.

Each third-party model should have a designated internal owner responsible for its governance, performance, and compliance. This role should be distinct from vendor relationship management and should sit within the institution's risk or control functions to preserve independence.

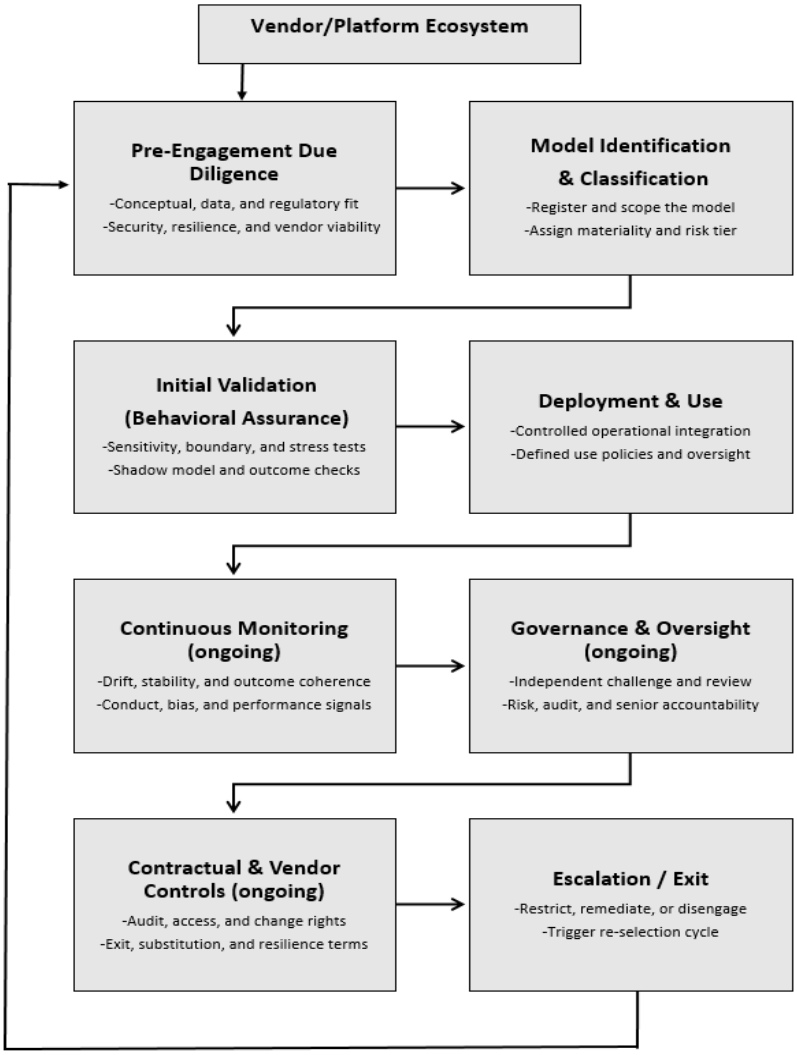

Fig 1: Lifecycle Governance of Third-Party Models

This figure illustrates the lifecycle governance of third-party models, showing how pre-engagement due diligence, model identification, behavioral validation, deployment, monitoring, governance oversight, contractual controls, and escalation or exit form a continuous control loop, with escalation feeding back into vendor re-selection.

4.1.4. Change and Release Management.

Institutions should establish processes to receive timely notice of model changes, assess their potential impact, and determine whether additional validation or controls are required. Even when institutions cannot prevent changes, they should be able to understand and respond to them.

4.1.5. Exit and Substitutability Planning.

Governance should include consideration of how models could be replaced or decommissioned if risks become unacceptable or relationships end. This reduces dependency and enhances operational resilience.

4.2. Integration with Risk Management and Governance Structures

Third-party model governance should be embedded within existing risk and governance structures rather than treated as a separate or purely technical function.

- Board and Senior Management Oversight: Senior management and the board should have visibility into the institution's reliance on third-party models, the associated risks, and the effectiveness of controls. This supports informed decision-making and accountability.
- Alignment with Enterprise Risk Management: Third-party model risk should be integrated into enterprise risk assessments, stress scenarios, and risk appetite statements. This ensures that model-related risks are considered alongside credit, market, operational, and other risks.
- Independent Oversight and Challenge: Independent risk, compliance, and audit functions should have the mandate and capability to review third-party

model governance, validation, and performance, and to escalate concerns where appropriate.

4.3. Validation and Monitoring

Traditional model validation assumes access to detailed model internals. For third-party models, validation must rely more heavily on indirect and outcome-focused techniques, supplemented by whatever transparency the vendor provides.

- Initial Validation: Before use, institutions should perform validation that assesses conceptual soundness, input-output behavior, performance on relevant data, and alignment with regulatory and business requirements.
- Ongoing Monitoring: Monitoring should include tracking performance metrics, stability indicators, drift measures, and business impacts over time. Monitoring thresholds should be defined to trigger review or escalation when deviations occur.
- Outcome Based Review: Institutions should periodically assess whether model outcomes remain appropriate, fair, and consistent with expectations, even if predictive performance remains high.

4.4. Auditability and Evidence

A central requirement of governance is the ability to demonstrate accountability to regulators and other stakeholders. Institutions should maintain documentation evidencing due diligence, validation activities, monitoring results, and decision-making processes related to third-party models. This documentation should be sufficient to support supervisory review and internal audit.

Audit functions should assess not only whether controls exist, but whether they are effective in practice. This includes reviewing how institutions respond to model changes, incidents, and emerging risks.

4.5. Contractual Enablement

Contracts are a critical enabler of governance in third-party settings. Without appropriate contractual rights, institutions may lack the ability to obtain information, perform oversight, or respond to issues.

Contracts should, where feasible, include provisions for:

- Audit And Inspection Rights;
- Access To Relevant Model Documentation And Performance Information;
- Timely Notification Of Material Changes And Incidents;
- Cooperation With Regulatory Inquiries; And
- Exit and Transition Arrangements.

These provisions do not eliminate risk, but they create the conditions under which governance, validation, and accountability can be exercised.

4.6. Proportionality and Practicality

Not all third-party models require the same level of governance. The framework emphasizes proportionality: governance intensity should reflect materiality, complexity, and potential harm. Overly burdensome controls can inhibit innovation, while insufficient controls can expose institutions to unacceptable risk. The objective is not to eliminate third-party model risk but to manage it in a way that is transparent, accountable, and consistent with regulatory expectations.

5. Validation Framework for Opaque and Black-Box Models

5.1. Context and Motivation

Independent model validation is a core requirement of model risk management frameworks globally. Its objective is to ensure that models are conceptually sound, empirically reliable, and appropriate for their intended use. Traditional validation practices assume that validators have access to a model's design documentation, theoretical foundations, implementation logic, and training data.

This assumption no longer holds in many contemporary financial environments. Increasingly, financial institutions rely on third-party vendors to provide embedded Artificial Intelligence (AI), Machine Learning (ML), and quantitative models through Software-as-a-Service (SaaS) and platform-based delivery models. These vendor models are often proprietary, continuously updated, and operationally controlled outside the institution. As a result, direct inspection of model internals is frequently impossible, while regulatory accountability for outcomes remains with the institution.

This structural mismatch between accountability and control creates a validation problem that is qualitatively different from traditional internal model validation. The purpose of this section is to develop a validation framework that addresses this problem directly by redefining validation as a process of behavioral assurance rather than internal inspection.

The objective is not to replicate traditional validation under constrained conditions, but to construct a disciplined, evidence-based alternative that enables institutions to demonstrate effective challenge, ongoing oversight, and regulatory accountability even when model transparency is limited. Recent academic work has emphasized the challenges of validating complex, opaque models using traditional inspection-based approaches, motivating a shift toward outcome- and behavior-based assurance [7].

5.2. Formal Setting and Observability Constraints

Let:

- $X_t \in \mathbb{R}^n$ denotes the input vector at time t.
- $f_t : \mathbb{R}^n \to \mathbb{R}$ denotes the (possibly changing) vendor model.
- $Y_t = f_t(X_t)$ denotes the model output.
- D_t denotes the joint distribution of (X_t, Y_t).
- θ_t denotes latent vendor parameters.

Opacity implies f_t and θ_t are unobservable; Institutions observe only samples of (X_t , Y_t), and possibly limited vendor-provided metadata. Validation therefore cannot be based on internal correctness of f_t . It must be based on observable properties of the mapping from inputs to outputs, and on how those properties evolve over time and across conditions.

5.3. Interpretation and Governance Implications of the Formal Diagnostics

The formal diagnostics introduced in this section are not abstract mathematical constructs, but operational tools that translate opaque model behavior into measurable, reviewable, and governable quantities. Each diagnostic corresponds to a specific dimension of model risk and supports a distinct aspect of governance, validation, and regulatory assurance.

The sensitivity vector S_t captures the responsiveness of model outputs to changes in individual inputs. It provides a quantitative representation of the model's internal logic as it is expressed through behavior, even when that logic cannot be directly inspected. Unexpected signs, magnitudes, or instability in sensitivity profiles indicate potential conceptual misalignment, overfitting, or fragility, and therefore serve as early warning indicators of model risk. From a governance perspective, high or unstable sensitivity suggests the need for increased monitoring, usage restrictions, or escalation.

The boundary instability measure I_t focuses on model behavior in extreme but plausible regions of the input space. Many material failures occur not in average conditions but near the edges of operational or economic regimes. By explicitly testing and measuring worst-case responsiveness, institutions can identify nonlinearities, cliffs, or unsafe regions that are invisible under normal operating conditions. Elevated boundary instability signals increased tail risk and informs both validation judgment and risk appetite decisions.

The shadow divergence metric Δ_t provides a structural point of comparison between the opaque vendor model and a transparent internal benchmark. Its purpose is not to assert that one model is superior, but to detect structural divergence over time. Persistent or increasing divergence can indicate regime shifts, hidden vendor model changes, or misalignment with institutional assumptions, thereby triggering investigation and challenge.

Drift decomposition separates observed change into distinct components input drift, concept drift, behavioral drift, and decision drift each of which has different causes and governance implications. This decomposition prevents the misdiagnosis of problems and enables targeted responses, such as data remediation, recalibration, policy adjustment, or usage review, rather than indiscriminate retraining or model replacement.

The outcome coherence metric C_t measures the stability of model impacts across segments and over time. It serves as a bridge between technical validation and conduct, fairness, and reputational risk considerations. Large or unexplained shifts in outcome distributions can signal emerging bias, unintended consequences, or structural changes that warrant review even if aggregate performance remains stable.

Finally, the evidence vector $V_t = (S_t , I_t , \Delta_t , D_t , C_t)$ integrates these diagnostics into a unified representation of model behavioral health. This vector functions as the central object of governance: it is what is reviewed by validation, monitored by risk, audited by internal audit, and presented to senior management or supervisors as evidence of ongoing oversight. Rather than relying on opaque vendor assurances or fragmented metrics, institutions can demonstrate disciplined, structured, and reviewable governance over externally controlled models.

Together, these diagnostics operationalize the concept of validation under opacity. They do not eliminate uncertainty, but they render it visible, measurable, and governable. This shift from unobservable internal correctness to observable behavioral assurance is the central methodological contribution of the proposed framework

5.4. Reframing Validation under Opacity: From Inspection to Behavioral Assurance

Under opacity, validation must shift from a model-centric paradigm ("What is inside the model?") to a behavior-centric paradigm ("How does the model behave across conditions, time, and populations?"). This reframing recognizes that models can be validated indirectly through their observable properties even when their internal mechanisms are inaccessible.

This reframing has three implications. First, validation becomes probabilistic rather than deterministic; it increases confidence but cannot prove correctness. Second, validation becomes continuous rather than episodic; assurance decays as models and environments change. Third, validation becomes multi-dimensional; assurance arises from consistency across independent lines of evidence rather than from any single test. These principles guide the design of the framework.

5.5. The Validation Assurance Ladder

The **Validation Assurance Ladder (VAL)** organizes validation activities into increasing levels of rigor, aligned with model materiality, opacity, and potential impact.

Table1: Validation Assurance Ladder (Val)

Level	Focus	Methods	Purpose
L1	Process assurance	Vendor documentation, governance review	Basic eligibility
L2	Behavioral coherence	Sensitivity and boundary testing	Detect anomalies
L3	Comparative integrity	Benchmarking and shadow models	Independent challenge
L4	Stress robustness	Regime and scenario testing	Resilience assessment

L5	Temporal stability	Drift and outcome monitoring	Ongoing assurance

Institutions should target higher levels for models with higher decision impact, customer impact, or systemic relevance.

5.6. Behavioral Testing

5.6.1. Sensitivity Analysis

Sensitivity analysis examines how outputs respond to controlled changes in inputs. Inputs are perturbed individually and in combinations to assess:

- Monotonicity (Do Outputs Move In Expected Directions?),
- Continuity (Are There Abrupt Jumps Or Cliffs?),
- And Proportionality (Are Changes Economically Plausible?).

Unexpected sensitivity patterns can indicate hidden dependencies, instability, or potential bias. Define the sensitivity vector: $S_t = \nabla_X [Y_t \mid X_t]$

Estimated via finite differences: $\hat{S}_{\{t,i\}} = \{ f_t (X_t + \delta e_i) - f_t (X_t) \} / \delta$

Sensitivity captures how strongly outputs respond to each input. Unexpected signs, magnitudes, or instability indicate conceptual misalignment, overfitting, or fragility.

Governance implication: High or unstable sensitivity warrants increased monitoring, usage restrictions, or escalation.

5.6.2. Boundary and Extreme Testing

Boundary testing evaluates model behavior at the edges of plausible input ranges. Extreme but realistic scenarios are used to identify:

- Numerical Instability,
- Implausible Outputs,
- Or Breakdowns In Decision Logic.

This is particularly important for regulatory capital, stress testing, and credit decisioning.

Define: $I_t = \sup_{\{x \epsilon \beta\}} | \partial f_t(x) / \partial x |$

Where β is the set of extreme but plausible inputs.

Boundary instability identifies nonlinearities, cliffs, and unsafe regions that are invisible in average conditions.

Governance implication: Elevated I_t indicates tail risk and informs risk appetite and resilience planning.

5.7. Benchmarking and Shadow Modeling

Benchmarking compares model outputs to:

- Alternative Vendor Models,
- Simpler Internal Models,
- Or Expert Judgment.

Shadow models are intentionally simpler and more transparent. Their purpose is not superior accuracy but interpretability and control. Persistent divergence between vendor and shadow models triggers investigation, not automatic rejection.

This comparative approach provides structural challenge without requiring internal access.

5.8. Regime and Scenario Stress Testing

Models are evaluated under simulated regime shifts such as:

- Rapid Interest Rate Changes,
- Liquidity Shocks,
- Economic Downturns,
- Or Portfolio Composition Shifts.

The objective is to assess whether model behavior remains stable, plausible, and aligned with institutional expectations under stress.

This mirrors financial stress testing logic applied to model behavior rather than balance sheets.

Let g be a transparent internal benchmark.

Define: $\Delta_t = [| f_t (X_t) - g (X_t)|]$

Persistent increases indicate structural divergence, regime shifts, or hidden vendor changes.

Governance implication: Rising Δ_t triggers investigation and vendor challenge.

5.9. Drift Decomposition

Rather than treating drift as a single phenomenon, the framework decomposes drift into:

- Input drift: changes in data distributions,
- Concept drift: changes in relationships between inputs and outcomes,
- Behavioral drift: changes in model output patterns,
- Decision drift: changes in how outputs are used operationally.

This decomposition enables targeted remediation rather than blanket recalibration.

Mathematically, Observed change is decomposed as:

$$D_t = D_{t-1} + \Delta_{\text{input}} + \Delta_{\text{concept}} + \Delta_{\text{behavior}} + \Delta_{\text{decision}}$$

Input drift = Jensen–Shannon divergence between $P_t (X)$ and $P_{\{t-1\}} (X)$

Behavioral drift = $\mathbb{E} [| Y_t - Y_{\{t-1\}} |]$

This prevents misdiagnosis and supports targeted remediation:

- Input drift → data review
- Concept drift → model review
- Behavioral drift → stability analysis
- Decision drift → policy or governance review

5.10. Outcome Coherence and Fairness

Validation extends beyond accuracy to include outcome coherence:

- Are outcomes consistent across comparable segments?
- Are changes explainable by business or economic shifts?
- Are there emerging disparate impacts?

This supports regulatory and ethical accountability.

Let outcome coherence be measured as the aggregate Wasserstein (W) distance between segment-level outcome distributions across time. Partition populations into segments k:

$$C_t = \Sigma_k \, \mathrm{W} \, \{P_t \, (Y|k), \, P_{\{t\text{-}1\}}(\, Y|k) \, \}$$

Outcome coherence measures whether impacts are shifting across segments.

Governance implication: Supports fairness, conduct, and reputational risk management.

5.11. Vendor Engagement

Validation is complemented by structured vendor engagement, including:

- review of change logs,
- testing summaries,
- incident reports,
- and governance practices.

Vendor input is treated as evidence, not assurance.

5.12. Evidence and Audit Trail

Validation must generate evidence that is:

- reproducible,
- traceable,
- explainable,
- and reviewable.

This includes structured reports, dashboards, issue logs, and escalation records.

Define:

$$V_t = (\, S_t \, , \, I_t \, , \, \Delta_t \, , \, D_t \, , \, C_t \,)$$

Validation holds if $V_t \in \forall$, where $\forall$, is the institution's acceptable region.

This vector becomes the central object of governance, review, audit, and supervisory communication.

5.13. Worked Examples and Regulatory Interpretation

This subsection illustrates how the proposed diagnostics operate in practice and how they support regulatory expectations such as effective challenge, ongoing monitoring, and proportionality.

Example 1: Sensitivity and Boundary Instability in a Credit Scoring Model

Assume a vendor credit model uses inputs:

- X_1 : income
- X_2 : debt-to-income ratio
- X_3: credit utilization

Suppose sensitivity estimates yield:

$S_t = (\hat{S}_{\{t,1\}} \, \hat{S}_{\{t,2\}} \, \hat{S}_{\{t,2\}}) = (0.01, -0.40, -0.05)$

Interpretation:

- Income has a weak positive effect (reasonable),
- Debt-to-income has a strong negative effect (reasonable),
- Credit utilization has moderate negative effect (reasonable).

Now boundary testing on high utilization reveals:

$I_t = \sup_{\{x \epsilon \beta\}} | \, \partial f_t(x) / \, \partial \mathrm{x} \, | = 3.5$

Interpretation:

A small change in utilization near the boundary causes large output changes, indicating instability in high-risk regions.

Governance implication: This supports effective challenge by identifying tail fragility even when average performance is stable. The institution may restrict use of the model for extreme cases or require enhanced monitoring.

Example 2: Shadow Model Divergence under Regime Change

Let g be a transparent logistic regression shadow model.

Suppose divergence evolves as:

Time	Δ_t
t_0	0.08
t_1	0.11
t_2	0.19

Interpretation:

Divergence is increasing, indicating that the vendor model is responding differently than the internal benchmark.

Governance implication: This triggers investigation possibly a vendor update or changing economic regime. This is a concrete form of effective challenge.

Example 3: Drift Decomposition

Suppose observed drift is decomposed as:

- Input drift: low (stable data distribution),
- Concept drift: high (error increases under same inputs),
- Behavioral drift: moderate,
- Decision drift: low.

Interpretation:

The model itself is changing or degrading, not the environment.

Governance implication: The institution challenges the vendor about retraining, model updates, or hidden changes.

Example 4: Outcome Coherence and Fairness

Partition borrowers into income deciles. Suppose:

$$C_t = \Sigma_k \, \text{Wasserstein} \, \{P_t \, (Y|k), \, P_{\{t\text{-}1\}}(\, Y|k) \, \} = 0.27$$

Interpretation:

Outcome distributions are shifting materially across segments.

Governance implication: Triggers fairness review and conduct risk assessment even if predictive accuracy remains stable.

5.14. Regulatory Alignment and Supervisory Interpretation

The validation framework proposed in this section is designed not merely as a technical methodology, but as an operational mechanism through which financial institutions can meet supervisory expectations in environments characterized by external control, limited transparency, and continuous model change. While regulatory frameworks emphasize accountability, effective challenge, ongoing monitoring, proportionality, auditability, and fairness, they provide limited guidance on how these principles should be operationalized when institutions cannot directly inspect or control model internals. This framework addresses that gap by translating regulatory principles into concrete, observable, and reviewable practices.

5.14.1. Effective challenge.

Supervisory guidance emphasizes that institutions must not rely uncritically on model outputs, but must exercise informed and independent judgment. In opaque third-party settings, traditional challenge based on code review or theoretical scrutiny is often unavailable. The framework therefore enables effective challenge through structured behavioral diagnostics. Sensitivity analysis S_t reveals how outputs respond to inputs and whether that behavior is conceptually coherent and stable. Shadow divergence Δ_t provides an independent point of comparison against transparent internal benchmarks. Drift decomposition D_t distinguishes between environmental change and model-driven change. Together, these mechanisms allow institutions to demonstrate that model behavior is actively interrogated, not passively accepted.

5.14.2. Ongoing monitoring.

Supervisory expectations require that validation extend beyond initial approval into continuous oversight. The framework operationalizes this requirement through systematic tracking of drift, stability, and outcome coherence over time. These measures allow institutions to detect degradation, regime shifts, or unintended consequences as they emerge, rather than retrospectively. Monitoring thus becomes an integral component of validation, enabling timely intervention and risk mitigation.

5.14.3. Proportionality.

Regulators expect governance and control intensity to be commensurate with model materiality, complexity, and potential harm. The Validation Assurance Ladder embeds this principle explicitly by linking the depth and rigor of validation activities to risk characteristics. This enables institutions to allocate resources proportionately while maintaining defensible oversight of high-impact systems.

5.14.4. Auditability and evidence.

A central challenge in third-party model governance is the ability to evidence compliance in the absence of internal artifacts. The framework addresses this by constructing an explicit evidence layer in the form of structured diagnostics, reports, dashboards, and escalation records, summarized in the validation evidence vector V_t This evidence is reproducible, traceable, and reviewable by independent functions and supervisors, enabling auditability even when model internals are inaccessible.

5.14.5. Fairness, conduct, and customer protection.

Regulators increasingly expect institutions to understand and manage the distributional and behavioral effects of automated decision systems. Outcome coherence analysis C_t provides a structured means of detecting shifting impacts across customer segments and over time. This supports fairness assessment, conduct risk monitoring, and customer protection objectives by ensuring that technical performance is evaluated alongside social and regulatory considerations.

5.14.6. Operational resilience and systemic stability.

Boundary instability testing I_t and regime stress testing extend validation into the domain of resilience by identifying failure modes, nonlinearities, and unsafe regions of model behavior. This helps prevent models from becoming hidden sources of fragility under stress or extreme conditions and supports broader resilience and stability objectives.

Taken together, these mechanisms enable institutions to demonstrate accountability in substance rather than merely in form. The framework allows institutions to convert abstract regulatory principles into operational practices and tangible evidence, reducing reliance on vendor assurances and strengthening the integrity, transparency, and resilience of model-driven decision-making. In doing so, it provides a practical bridge between supervisory expectations and the realities of platform-based financial infrastructure.

5.15. Limits and Residual Risk

This framework does not eliminate uncertainty. It structures uncertainty so it can be governed, monitored, and communicated. Residual risk remains and must be managed through governance, capital, and contingency planning.

5.16. Contribution of the Framework

This section contributes a formalized validation framework for opaque, externally controlled models that integrates behavioral diagnostics, comparative challenge, drift analysis, and governance escalation into a unified, audit-grade structure. It extends traditional validation into environments where transparency cannot be assumed and accountability cannot be outsourced.

The proposed framework does not attempt to recreate internal transparency. It constructs a disciplined substitute: structured behavioral evidence, comparative challenge, and continuous monitoring that together provide defensible assurance under opacity.This enables institutions to manage third-party model risk not through blind trust, but through measurable, reviewable, and enforceable controls.

While prior literature and regulatory guidance have discussed aspects of model risk, outsourcing risk, and AI

governance separately, this paper is among the first to integrate technical validation, legal enforceability, and organizational governance into a unified framework specifically designed for opaque, externally controlled models in regulated financial environments.

6. Contractual Controls and Legal Enablement of Third-Party Model Governance

6.1. The Role of Contracts in Model Risk Governance

In traditional internal model environments, governance and validation are implemented primarily through organizational mechanisms such as policies, procedures, independent review, and internal accountability structures. In third-party environments, these mechanisms remain necessary but are no longer sufficient. When models are developed, maintained, and updated outside the institution, the practical ability to govern and challenge them depends critically on the legal rights and obligations defined in contractual arrangements.

Contracts therefore function not merely as commercial instruments but as elements of governance infrastructure. They determine whether institutions can obtain information, exercise oversight, escalate concerns, respond to incidents, and disengage from relationships when risks become unacceptable. Without appropriate contractual enablement, even robust internal governance frameworks may lack the practical authority needed to be effective.

This section develops a framework for contractual controls that translates governance and validation requirements into enforceable legal mechanisms, thereby aligning legal, technical, and regulatory dimensions of third-party model risk management.

6.2. Extending Third-Party Risk Management to Model Risk

Traditional third-party risk management has focused on operational continuity, information security, financial stability of vendors, and regulatory compliance. While these concerns remain important, they do not fully capture the risks associated with embedded models that directly influence financial decisions, customer outcomes, and regulatory metrics.

Model risk introduces additional requirements that are not adequately addressed by standard vendor management approaches. Institutions require not only service availability and data protection, but also visibility into model behavior, awareness of model changes, the ability to perform independent assessment, cooperation during regulatory reviews, and the ability to exit or substitute when risks cannot be mitigated.

As a result, contracts must evolve from general vendor risk instruments into specific model risk enablement mechanisms that support accountable and defensible model governance.

Several industry initiatives have proposed governance and contractual principles for AI and third-party systems (FINOS, 2024), reflecting growing awareness of these risks, although such initiatives typically remain high-level and do not address validation under opacity or regulatory alignment in a systematic way.

6.3. Contractual Domains Supporting Model Governance

Contractual provisions relevant to model risk governance can be grouped into several interrelated domains. These domains do not operate in isolation; together they form the legal foundation that enables technical and organizational controls to function.

Transparency and information rights establish the institution's ability to understand and govern externally developed models. While full disclosure of proprietary intellectual property is rarely feasible, institutions can reasonably require access to high-level descriptions of model purpose, inputs, limitations, development practices, and known constraints. Such information is necessary to support responsible use, internal validation, and regulatory explanation.

Change management and notification provisions address the dynamic nature of vendor-managed models. Institutions must be informed of material changes that may affect model behavior, performance, or regulatory relevance. This enables alignment between validation activities, operational use, and governance oversight, and prevents silent model evolution from undermining prior assurances.

Audit, inspection, and evidence rights support accountability and review. Institutions must be able to obtain assurance that appropriate controls exist and are functioning, either through direct audit rights or through access to independent assurance reports. These mechanisms provide the evidentiary basis for internal audit, supervisory engagement, and governance escalation.

Incident management and remediation clauses define how failures, anomalies, or adverse events are handled. Clear expectations regarding incident notification, investigation, root cause analysis, and remediation ensure that model failures are treated as governed risk events rather than isolated technical issues.

Regulatory cooperation provisions support supervisory engagement. Institutions are accountable to regulators for the outcomes of models they use, regardless of vendor involvement. Contracts therefore must support information sharing, regulatory access, and vendor cooperation during examinations, subject to appropriate confidentiality protections.

Exit and substitutability provisions reduce dependency and concentration risk. Institutions must retain the ability to disengage from relationships that pose unacceptable risk and to transition to alternative solutions where necessary. This

supports resilience and reduces the risk of lock-in undermining governance.

6.4. Contracts as Risk Mitigation Instruments

When properly designed, contractual controls function as active risk mitigants rather than passive legal safeguards. They enable institutions to convert abstract governance expectations into enforceable obligations, thereby reducing reliance on trust, goodwill, or informal assurances. In this way, contracts become part of the risk control system itself, complementing technical validation, monitoring, and organizational oversight.

This integration also supports internal accountability. When legal rights and obligations are aligned with governance responsibilities, internal stakeholders are empowered to act when risks arise, and escalation pathways become credible rather than symbolic.

6.5. Governance Integration and Review

Contractual controls should not be designed or managed in isolation. They should be developed jointly by legal, risk, compliance, procurement, and business stakeholders, and aligned with the institution's model risk management framework and risk appetite. As technologies, regulatory expectations, and business strategies evolve, contracts must be reviewed and updated to ensure continued relevance and effectiveness.

Internal audit plays a key role in assessing whether contractual controls are adequate, implemented, and effective in practice. This reinforces the treatment of contracts as living governance instruments rather than static legal documents.

6.6. Limits and Practical Constraints

Not all contractual rights are achievable in all contexts. Vendors may resist transparency, audit, or regulatory access on intellectual property or commercial grounds. Institutions must therefore balance risk reduction against feasibility and market realities, recognizing that contractual controls reduce but do not eliminate risk.

The objective is not contractual perfection, but sufficient legal enablement to support accountable governance and regulatory defensibility.

6.7. Contribution

This section contributes a structured framework for translating model governance requirements into enforceable legal mechanisms. By treating contracts as components of the risk system rather than as separate commercial artifacts, it extends model risk management into the legal and institutional domain, addressing a critical gap in existing practice and enabling more robust governance of third-party models.

7. Systemic Risk, Concentration, and the Platformization of Financial Models

7.1. From Institutional Risk to Systemic Exposure

Model risk has traditionally been treated as an institution-specific concern, arising from errors in design, implementation, or use of internal analytical tools. However, the structural transformation of financial infrastructure toward platform-based delivery models has altered the scale and transmission of model risk. When large numbers of institutions rely on common vendors, shared platforms, or standardized analytical components, model behavior and model failure cease to be isolated events and become potential sources of correlated and systemic exposure.

This shift does not imply that all third-party models pose systemic risk, nor that platformization is inherently destabilizing. Rather, it changes the conditions under which localized weaknesses can propagate across institutions, markets, and time horizons. Understanding this transition is essential for both institutional governance and macro-prudential oversight.

7.2. Mechanisms of Risk Propagation

Platform-based models create several mechanisms through which risks can propagate beyond individual institutions.

First, common dependency introduces correlation. When multiple institutions use the same or similar vendor models trained on similar data and optimized for similar objectives, their decisions may become synchronized. This synchronization can amplify market movements, reinforce pro-cyclical behavior, or generate clustering of exposures that is not visible at the level of any single institution.

Second, opacity limits early detection. If vendor models are proprietary and their internal behavior is not transparent, institutions may detect emerging weaknesses only through observable outcomes. If many institutions observe similar anomalies at the same time, corrective action may be delayed and collective, increasing the risk of abrupt adjustments.

Third, continuous and centralized change can act as a transmission channel. Vendor-driven updates, retraining, or feature changes can affect multiple institutions simultaneously. Even benign changes may have heterogeneous effects across portfolios, markets, or regulatory contexts, potentially introducing correlated shocks.

Fourth, concentration creates dependency. Heavy reliance on a small number of dominant providers reduces substitutability and increases the potential impact of vendor-specific failures, whether technical, operational, legal, or financial.

These mechanisms do not imply inevitability of systemic harm, but they do alter the topology of risk transmission in ways that traditional institution-centric frameworks do not fully capture.

7.3. Implications for Institutional Governance

From an institutional perspective, these dynamics expand the scope of model risk management. Governance must account not only for how a model behaves within a specific portfolio or business line, but also for how that model's behavior may interact with broader market dynamics and shared dependencies.

This does not require institutions to model the entire financial system. It does require awareness of dependency structures, participation in industry dialogue, and engagement with supervisory initiatives aimed at understanding concentration and common exposures.

Institutions should therefore treat concentration and dependency as explicit risk factors in vendor selection, governance intensity, and exit planning. Diversification of providers, development of internal contingency capabilities, and contractual provisions supporting transition and substitutability are practical mechanisms for reducing systemic vulnerability at the institutional level.

7.4. Supervisory and Macro-Prudential Considerations

From a supervisory perspective, platform-based model risk challenges traditional regulatory boundaries. Supervisors typically oversee institutions, not vendors, and they assess risk primarily within institutional balance sheets. Platformization introduces cross-institutional linkages that are not easily captured by firm-level supervision alone.

This does not necessitate a fundamental redesign of regulatory frameworks, but it does suggest a greater role for horizontal reviews, thematic examinations, and information sharing across institutions. Supervisors may also increasingly focus on critical third-party providers as part of broader operational resilience and systemic risk initiatives.

The framework developed in this paper complements such efforts by enabling institutions to generate structured evidence about model behavior, dependencies, and changes. This evidence can support supervisory dialogue, facilitate early identification of emerging risks, and contribute to a more informed macro-prudential perspective.

7.5. The Role of Governance in Mitigating Systemic Risk

While systemic risk cannot be eliminated at the institutional level, governance can mitigate its formation and amplification. By embedding behavioral monitoring, drift detection, contractual controls, and exit planning into model governance, institutions reduce the likelihood that weaknesses remain hidden, unmanaged, or unaddressed.

Moreover, when institutions adopt similar disciplined governance practices, collective resilience increases. Transparency, challenge, and accountability at the micro level support stability at the macro level by reducing the probability of synchronized failures and uncontrolled propagation.

In this sense, robust third-party model governance is not only a matter of institutional prudence, but also a contribution to financial system stability.

7.6. Contribution

This section extends the analysis of third-party model risk beyond the institutional boundary, articulating how platform-based delivery models alter the structure and transmission of risk across the financial system. By linking micro-level governance mechanisms to macro-level stability considerations, it highlights the broader significance of third-party model governance and reinforces its relevance to supervisors, policymakers, and the financial system as a whole.

8. Conclusion

The increasing reliance of financial institutions on externally developed and operated analytical systems represents a fundamental shift in how model risk is created, transmitted, and governed. As Artificial Intelligence, Machine Learning, and quantitative models become embedded within platform-based infrastructures, traditional assumptions about transparency, control, and institutional self-sufficiency no longer hold. At the same time, regulatory accountability for model outcomes remains firmly with institutions. This structural tension between responsibility and control defines the contemporary challenge of third-party model risk.

This paper has addressed that challenge by developing an integrated framework for the governance, validation, and legal enablement of third-party models embedded within SaaS and Risk-as-a-Service platforms. It has argued that third-party model risk cannot be treated simply as a subset of either internal model risk or traditional vendor risk, but constitutes a distinct configuration of risk characterized by external control, opacity, continuous change, and concentration.

The paper's first contribution is the articulation of this risk configuration and its implications for governance. By identifying the structural features that differentiate third-party models from internal ones, the paper provides a conceptual foundation for tailored oversight rather than the mechanical extension of existing frameworks.

The second contribution is the development of a validation framework designed explicitly for opaque and externally controlled models. By reframing validation as a process of behavioral assurance rather than internal inspection, and by formalizing diagnostics for sensitivity, stability, drift, and outcome coherence, the framework enables institutions to exercise effective challenge and ongoing oversight even when internal model artifacts are unavailable. This extends the scope of model risk management into environments where traditional validation approaches are infeasible.

The third contribution is the integration of legal and contractual controls into the model risk framework. By

treating contracts as governance instruments rather than purely commercial documents, the paper shows how legal rights and obligations can enable, rather than merely constrain, technical and organizational controls. This integration addresses a critical gap in existing practice and supports accountable governance in third-party contexts.

Finally, the paper situates third-party model governance within a broader systemic context. Platform-based delivery models alter the topology of risk transmission by creating shared dependencies, synchronized behavior, and concentration. Robust micro-level governance is therefore not only an institutional necessity, but also a contributor to financial system stability.

Together, these contributions do not propose a single prescriptive solution, nor do they claim to eliminate the risks inherent in third-party models. Rather, they offer a structured, defensible, and adaptable approach to managing those risks in a way that aligns regulatory expectations with technological and institutional realities.

As financial infrastructures continue to evolve, the governance of externally embedded models will remain a dynamic and contested space. The frameworks proposed in this paper are intended as a foundation for ongoing development rather than as a final answer. Future work may refine these mechanisms, extend them to new domains, and integrate them with emerging supervisory approaches. What remains constant is the need for disciplined accountability, informed challenge, and institutional responsibility in an increasingly interconnected and platform-driven financial system.

References

[1] Board of Governors of the Federal Reserve System (2011). *SR 11-7.*

[2] Prudential Regulation Authority (2023). *SS1/23.*

[3] European Union (2022). *Digital Operational Resilience Act (DORA).*

[4] Monetary Authority of Singapore (2023). *Technology Risk Management / Outsourcing.*

[5] Financial Stability Board (2017). *AI and ML in Financial Services.*

[6] European Central Bank (2024). *Implications of AI for financial stability and cyber risk.*

[7] British Actuarial Journal (2024). *Model Risk: Illuminating the Black Box.*

[8] P1. FINOS (2024). *AI Governance Framework — Legal and Contractual Controls.*