INTERNATIONAL JOURNAL OF ACTA INFORMATICA
VOLUME (2025)

MODEL RISK MANAGEMENT FOR ADAPTIVE CREDIT EARLY-WARNING SYSTEMS IN SUPERVISED SMALL-BUSINESS LENDING: ALIGNING MACHINE-LEARNING RECALIBRATION WITH FEDERAL MODEL-RISK GUIDANCE AND THE NIST AI RISK MANAGEMENT FRAMEWORK

Naima Bintay Karim[1*], Tomasz Kovac[2], Zhu Song-Chun[3]
[1]Arkansas State University, USA
[2] Warsaw Applied AI Laboratory, POLAND
[3] Tsinghua University, CHINA

*Corresponding Author Email: naimabintay980@gmail.com

ABSTRACT

SR 11-7, issued jointly by the Federal Reserve and OCC in April 2011, remains a foundational U.S. standard for model risk management. Its framework is based on three core pillars: sound model development, effective independent validation and challenge, and strong governance. Although the guidance broadly applies to banking organizations supervised by the Federal Reserve and OCC, the FDIC extended it in 2017 to FDIC-supervised institutions with $1 billion or more in assets. The depth of model validation is expected to reflect a model's materiality, complexity, and risk. However, SR 11-7 was developed before modern machine-learning systems and continuously recalibrated models became widespread. Consequently, it provides limited operational guidance for adaptive credit early-warning models that are frequently updated using new data. This creates particular governance challenges for community banks, credit unions, and Community Development Financial Institutions (CDFIs). The article therefore examines how the NIST AI Risk Management Framework (AI RMF), published in January 2023 as a voluntary framework, can complement SR 11-7. AI RMF organizes AI risk management into four functions: Govern, Map, Measure, and Manage. These functions can provide additional AI-specific governance detail where SR 11-7 is less explicit. The proposed approach combines SR 11-7 and NIST AI RMF to establish stronger governance for adaptive credit early-warning systems. It emphasizes comprehensive model documentation, evidence supporting recalibration, independent validation, governance controls, and continuous audit trails, ensuring that dynamically updated models remain explainable, controlled, and appropriately validated.

KEYWORDS: SR 11-7; NIST AI RMF; Model Validation; Community Banks; Regulatory Compliance; Model Governance

INTRODUCTION

This article draws on the foundational federal interagency model risk management

guidance, the NIST AI Risk Management Framework, and supervisory commentary on their application to community financial institutions, to address a governance question with direct, practical urgency for any institution deploying an adaptive credit early-warning model: how should model risk management be structured for a system that, unlike a traditional static credit scorecard, is designed to recalibrate continuously as new data and fraud or default patterns emerge.

This article is organised to move from regulatory history to practical application. Sections 1 and 2 establish what SR 11-7 requires and why its principles matter for institutions currently operating adaptive early-warning models. Section 3 introduces the NIST AI RMF as a complementary standard and maps its four functions to the specific governance questions such a model raises. Sections 4 through 8 address the recalibration-specific, forward-looking, institution-tier, CDFI-specific, and documentation-architecture considerations that translate both frameworks into practice. Sections 9 and 10 address limitations and examination readiness directly, and Sections 11 and 12 close with recommendations and conclusions grounded in the evidence assembled throughout.

This question has particular salience given how much the model landscape has evolved since SR 11-7's original 2011 issuance. Federal model risk guidance has remained substantively stable for more than a decade: the Federal Reserve and OCC jointly issued SR 11-7 and OCC Bulletin 2011-12 in April 2011, the FDIC extended equivalent guidance to FDIC-supervised institutions with $1 billion or more in assets via FIL-22-2017 in 2017, and the agencies extended model risk principles to BSA/AML transaction-monitoring and sanctions-screening systems through an Interagency Statement in 2021 [1][7][8]. Table 1 and Figure 1 set out this regulatory timeline.

Table 1. Timeline of U.S. federal model risk management guidance, 2011–2024.

Date	Event	Significance
2011	Federal Reserve SR 11-7 and OCC Bulletin 2011-12 issued	Establishes the three-pillar model risk management framework (development, effective challenge, governance) that remains in force today [1][7]
2017	FDIC adopts equivalent guidance via FIL-22-2017	Extends coverage to FDIC-supervised institutions with $1 billion or more in assets [8]
2021	Interagency Statement on BSA/AML Model Risk Management	Extends model risk principles to transaction monitoring and sanctions screening systems [8]
Jan 2023	NIST publishes AI Risk Management Framework	Voluntary, sector-agnostic AI governance framework organised around

	1.0	Govern/Map/Measure/Manage [2][3]
2023–2024	Growing industry commentary on applying SR 11-7 to AI/ML models	Confirms adaptive, machine-learning-based models fall within SR 11-7's model definition even though the guidance predates modern ML practice [3][7]

Figure 1. Timeline of U.S. federal model risk management guidance, 2011–2024 [1][2][3][7][8].

Many community banks read SR 11-7's language as an effectively binding annual validation requirement, in some cases building out-of-scale validation teams as a result, even though the guidance's own proportionality language does not strictly compel a uniform annual cadence regardless of model risk [7][8]. Applying SR 11-7's principles thoughtfully, rather than defaulting to the most conservative possible reading, requires a genuine, ongoing exercise of judgement, particularly for the adaptive, machine-learning-based early-warning systems addressed throughout this article series.

THE THREE PILLARS OF SR 11-7 AND WHAT THEY REQUIRE

SR 11-7 is organised around three core pillars — sound model development, effective independent challenge, and governance — and expects the depth and frequency of validation activity to scale with a model's materiality, complexity, and risk contribution, even though the guidance does not formalise this scaling into named, numbered tiers the way some more recent international and industry frameworks do [7][8]. Table 2 summarises how each pillar applies specifically to an adaptive credit early-warning model.

Table 2. SR 11-7's three pillars applied to adaptive credit early-warning models.

Pillar	Core requirement	Application to an adaptive early-

	under SR 11-7	warning model
Sound development, implementation, and use	Rigorous development process, conceptual soundness, testing, and documentation [7]	Extends to every recalibration event, not just initial deployment, since each recalibration is, in effect, a new model version
Effective challenge (independent validation)	Independent review by qualified staff free from developer influence, with depth proportional to materiality and complexity [7][8]	Requires validation resourcing that keeps pace with recalibration frequency, not a single point-in-time review
AI/ML scope	Silent on AI/ML specifically at the time of 2011 issuance; subsequent industry and supervisory practice confirms AI/ML models fall within SR 11-7's broad model definition [3][7]	An adaptive, continuously recalibrated but fundamentally predictive (rather than generative) early-warning model falls within this definition and is therefore in scope, even though a generative-AI component in the same workflow (for example, an LLM-drafted explanation of a flag) sits in a governance grey area SR 11-7 itself does not address
Governance, policies, and controls	Board/senior-management oversight, a comprehensive model inventory, and clear roles and responsibilities [7][8]	Model inventory entry must capture the model's adaptive nature and recalibration cadence, not just its initial specification

The proportionality principle embedded in SR 11-7 has real teeth for community institutions, even without a formalised, numbered tier system. A model with high inherent risk, broad exposure, or a consequential purpose warrants the fullest level of lifecycle oversight the guidance describes, while a lower-materiality model can reasonably receive proportionately lighter validation, provided the institution can evidence that judgement with documented rationale rather than simply defaulting to a uniform cycle regardless of risk [7][8]. An adaptive credit early-warning model that feeds loan-restructuring recommendations warrants this fullest level of oversight at most institutions, given that its output directly influences credit availability for individual borrowers — precisely the kind of consequential, fair-lending-relevant use case SR 11-7's proportionality principle is designed to flag for the most rigorous validation and governance, independent of the deploying institution's overall asset size.

A second consideration concerns AI specifically. SR 11-7 defines a model broadly as involving quantitative methods, systems, or approaches that apply statistical, economic,

financial, or mathematical theories and assumptions to process input data into quantitative estimates — a definition written in 2011 without generative AI, agentic AI, or modern machine learning specifically in view, though subsequent supervisory practice and industry commentary treat AI/ML systems used for quantitative estimation as falling within this definition [3][7]. An adaptive, continuously recalibrated but fundamentally predictive (rather than generative) early-warning model of the kind addressed throughout this article series falls within SR 11-7's definition of a model and is therefore in scope, but institutions building generative-AI-assisted components into the same workflow — for example, an LLM-drafted explanation of a flagged account, addressed in the responsible-AI-governance article in this series — face a genuine interpretive question, since SR 11-7 itself offers no explicit treatment of generative or agentic AI and institutions must currently apply broader risk-management judgement to such components [3].

This interpretive gap is worth dwelling on because it creates a genuine governance seam that institutions need to actively manage rather than assume away. A single end-to-end early-warning and loan-restructuring workflow of the kind described in the integration-architecture article in this series may combine a predictive scoring model (clearly within SR 11-7's model definition) with a generative-AI component that drafts a plain-language explanation of the flag for the loan officer (a component SR 11-7's 2011 text simply did not contemplate). Treating the whole workflow as governed only by whatever formal standard most obviously applies to its most conventional component risks under-governing the generative piece; waiting for dedicated generative-AI supervisory guidance before governing it at all risks leaving a consequential system component ungoverned indefinitely. The more defensible approach, consistent with both frameworks reviewed in this article, is to govern each component according to the most relevant available standard — SR 11-7 for the predictive model, the institution's broader AI governance practices informed by the NIST AI RMF for the generative component — while still tracking the combined workflow's overall risk at the Govern-function level described in Section 3.

THE NIST AI RISK MANAGEMENT FRAMEWORK AS A COMPLEMENTARY STANDARD

Where SR 11-7 provides the supervisory floor for model risk management at a regulated depository institution, the NIST AI Risk Management Framework provides a more AI-specific, voluntary complement that fills governance detail SR 11-7 itself does not fully specify. Published in January 2023 as NIST AI 100-1, the AI RMF is organised around four functions — Govern, Map, Measure, and Manage — developed through a public consultation process that drew approximately 400 sets of comments from more than 240 organisations [2][9]. Table 3 maps these four functions to the specific governance

questions an adaptive credit early-warning model raises.

Table 3. NIST AI RMF functions applied to an adaptive credit early-warning model.

NIST AI RMF function	Core question	Application to an adaptive credit early-warning model
Govern	Who is accountable for this system's risk?	Assigns model ownership, recalibration-approval authority, and escalation paths [2][10]
Map	Where does this system influence high-stakes decisions?	Identifies that model output feeds loan-restructuring recommendations, a fair-lending-relevant decision [3][10]
Measure	How do we quantify this system's risk?	Tracks precision/recall drift, feature drift, and demographic-parity metrics over time [3]
Manage	How do we respond when risk is detected?	Triggers recalibration, human review, or model suspension per pre-defined thresholds [3]

The AI RMF's Govern function is explicitly cross-cutting — it sits across the whole framework as the organisational foundation of policies, accountability, and culture that make the other three functions possible, rather than a discrete first step completed once and left behind [9]. This design principle maps directly onto the SR 11-7 governance pillar discussed in Section 2: an institution cannot credibly document a proportionate validation and oversight posture without first establishing exactly the kind of clear ownership and accountability structure the Govern function requires. In practice, this means an institution's SR 11-7 model-inventory and governance documentation and its NIST AI RMF governance documentation should be built as a single, integrated artifact rather than two separate compliance exercises addressing the same underlying accountability question from different regulatory vocabularies.

The NIST framework's development process also lends it interpretive weight beyond its voluntary status: it was built through a public consultation drawing roughly 400 comment submissions from more than 240 organisations, giving it a breadth of practitioner input that a purely internal institutional framework would lack [9]. For a community institution without the in-house AI governance expertise a large bank might have, this means the AI RMF Playbook's suggested actions are not merely one vendor's or consultant's opinion of good practice, but a synthesis of cross-industry input specifically designed to be adapted to an institution's own context — which is precisely the adaptation this article undertakes in Table 3 for the credit early-warning use case.

The core AI RMF document runs to roughly 40 pages, but NIST's companion Playbook — offering suggested actions, transparency questions, and reference resources for each function — runs to more than 140 pages, reflecting the framework's design intent that

the core document provides the abstractions while the Playbook carries the operational detail institutions actually need to implement it [9]. Figure 2 shows this relative scale.

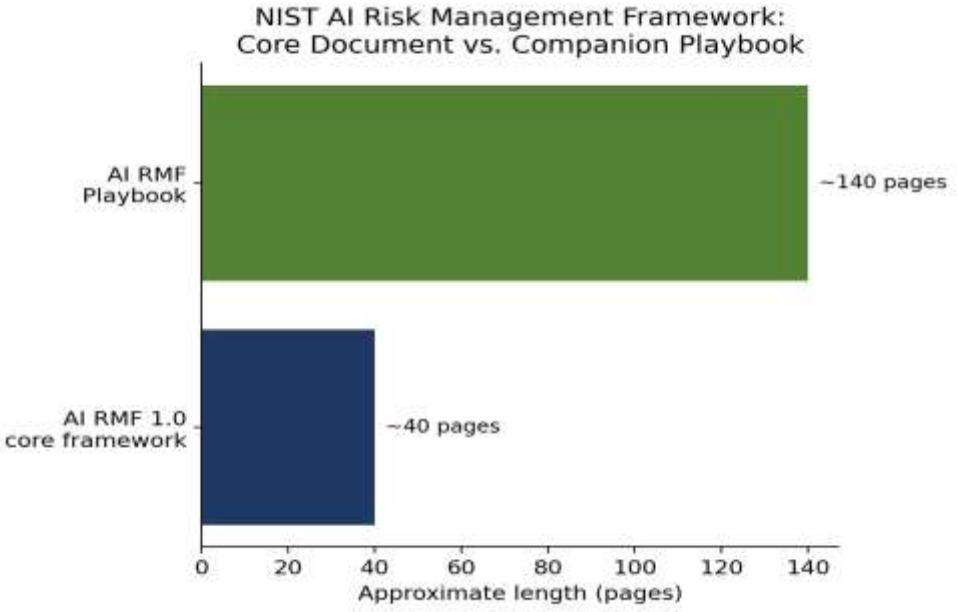

Figure 2. NIST AI Risk Management Framework: core document vs. companion Playbook, approximate length [9].

For an institution building model risk documentation for an adaptive early-warning system, the Playbook is the more directly useful artifact of the two: its function-by-function suggested actions can be mapped one-to-one against the recalibration-monitoring framework addressed in Section 4, giving an institution a documented, defensible basis for each control it puts in place, rather than needing to invent a monitoring framework from first principles.

RECALIBRATION-SPECIFIC MODEL RISK CONSIDERATIONS

An adaptive, continuously recalibrated early-warning model raises model risk management questions a traditional, infrequently-updated credit scorecard does not, and neither SR 11-7 nor the AI RMF addresses continuous recalibration in the level of operational detail community-institution risk teams need, since both predate widespread adoption of continuously updated ML systems in community lending. Three considerations follow from the principles both frameworks do establish. First, SR 11-7's proportionality principle (Section 2) cuts both ways: a model that recalibrates frequently arguably requires more frequent — not less — outcomes analysis and monitoring than the proportionality principle would suggest for a static model of comparable materiality, since each recalibration event is, in effect, a new model version requiring its own validation evidence. Second, the AI RMF's Measure function requires

quantitative and qualitative risk assessment on an ongoing basis, which for a recalibrating model means tracking not just point-in-time accuracy but the trajectory of accuracy, precision, and recall across each recalibration cycle, so that a gradual, cumulative drift across many small recalibrations — each individually defensible — does not go undetected simply because no single recalibration event looked alarming in isolation. Third, both frameworks' emphasis on documented, evidenced decision-making means every recalibration event needs a preserved record of the before/after performance comparison that justified it, precisely the audit-trail requirement addressed in the integration-architecture article in this series.

This recalibration-specific documentation burden is a genuine, non-trivial addition to standard model risk practice, and institutions should budget for it explicitly rather than assuming their existing static-model validation processes transfer directly. A validation team accustomed to an annual or biennial review cycle for a traditional scorecard needs a materially different, more continuous engagement model for a system that may recalibrate monthly or even more frequently, which has direct staffing and tooling implications.

A useful practical benchmark for calibrating this staffing implication comes from the model-performance literature reviewed elsewhere in this article series: published fraud- and credit-risk-detection studies commonly report ensemble models achieving accuracy in the 0.91–0.94 range, figures that are only meaningful when read alongside the specific validation date and data vintage they were measured against. An institution's recalibration log should preserve this same level of specificity for its own model — not simply "the model was recalibrated in March" but the specific before/after precision, recall, and any relevant fairness-metric comparison, dated and tied to the specific data snapshot used — so that a reviewer six months or two years later can reconstruct exactly what changed and why, without needing to rely on institutional memory that, per the broader legacy-system literature discussed in the integration-architecture article in this series, has a way of degrading exactly when it is most needed.

FUTURE DIRECTIONS

Three developments are likely to shape this regulatory landscape over the coming years. First, industry commentary increasingly anticipates that federal regulators will eventually need to address generative and agentic AI governance more explicitly, whether within an updated version of SR 11-7 or through separate guidance, given how far the model landscape has evolved since 2011; institutions should expect the governance seam discussed in Section 2 to eventually close rather than persist indefinitely, even though no specific timeline for such action has been announced. Second, as more institutions gain practical experience applying SR 11-7's proportionality principle to modern AI/ML systems, expect industry-standard tiering

rubrics to emerge — potentially through trade associations, examiner guidance, or shared vendor frameworks — that reduce the current reliance on each institution independently interpreting materiality for itself, a gap the illustrative framework in Table 4 is intended to help bridge in the interim. Third, the NIST AI RMF itself is expected to continue evolving with additional guidance and expanded profiles over time, and a sector-specific credit-lending profile, comparable to existing NIST profiles for other domains, would materially reduce the institution-specific adaptation work this article undertakes in Table 3 if and when NIST publishes one.

APPLYING THE COMBINED FRAMEWORK BY INSTITUTION TIER

Table 4 illustrates how the combined SR 11-7/NIST AI RMF approach might scale across three representative community-institution asset tiers, recognising that SR 11-7 itself is principles-based and does not draw a hard scope line at a specific asset threshold the way FDIC's FIL-22-2017 extension does at $1 billion, so even an institution below that threshold should apply proportionate, evidenced validation rather than concluding it is simply out of scope [7][8].

Table 4. Illustrative application of combined model risk governance by institution asset tier.

Institution tier	SR 11-7 posture	NIST AI RMF application
Small community bank/credit union (<$1B)	Document a high-materiality classification for the early-warning model given its consequential use, even though below FDIC's FIL-22-2017 $1 billion threshold, with proportionate (not maximal) validation depth	Apply Govern/Map at full depth; scale Measure/Manage tooling to available staff capacity
Mid-sized institution ($1B and above, FDIC FIL-22-2017 threshold)	Full three-pillar validation given FDIC's explicit coverage at this threshold; documented rationale for any reduced-intensity treatment elsewhere	Apply all four functions proportionately; Measure/Manage automation as volume warrants
Large, complex institution	Full SR 11-7 applicability with the most extensive validation and governance infrastructure, consistent with the guidance's risk-based intent [7]	Full four-function AI RMF implementation with dedicated AI governance staff

Even the smallest institution tier in Table 4 should not read "proportionate" as

"minimal." SR 11-7's own text sets a consistently high bar regardless of institution size: examiners reviewing a community bank's program will look for a complete model inventory with documented risk assessment, validation completed and current for high-materiality models, documented rationale for any lower-intensity validation elsewhere, demonstrated effective challenge through genuine independence, and clear board/senior-management oversight [7][8].

PRACTICAL IMPLICATIONS FOR CDFIS AND MISSION-DRIVEN LENDERS

Community Development Financial Institutions warrant separate consideration given their frequently smaller scale and mission-driven, higher-risk-tolerance underwriting approach discussed elsewhere in this series. A CDFI is highly unlikely to approach FDIC's $1 billion FIL-22-2017 threshold, but the principles-based rather than purely scope-based nature of SR 11-7 means a CDFI deploying an adaptive early-warning model that materially influences credit-restructuring decisions for underserved borrowers should still apply the high-materiality documentation discipline described in Section 2 — arguably with heightened rather than reduced attention to the fair-lending dimension of the Map function in Table 3, given that CDFI borrower populations are disproportionately represented in the protected-characteristic-correlated categories that make model-driven credit decisions most consequential to get right.

This heightened attention has a concrete governance implication: where a mid-sized bank might reasonably document a lower-materiality classification for a model with more limited borrower-consequence per the illustrative framework in Table 4, a CDFI's underlying mission of extending credit specifically to borrowers conventional lenders decline means an equivalent early-warning model is more likely to warrant the fullest validation treatment even at a much smaller asset scale, because the consequence of a false-positive flag falls disproportionately on exactly the borrower population the institution exists to serve. Regulators reviewing a CDFI's model risk practices, whether through direct prudential supervision (for bank- or credit-union-type CDFIs) or through CDFI Fund compliance review (for non-depository CDFIs), are likely to weigh this mission-alignment consideration alongside the more conventional materiality factors SR 11-7 and the NIST AI RMF were primarily designed around.

A UNIFIED DOCUMENTATION ARCHITECTURE

Sections 2 through 6 establish that SR 11-7 and the NIST AI RMF, while issued by different bodies for different purposes, converge on a common set of underlying documentation needs for an adaptive credit early-warning model. Table 5 makes this convergence explicit, mapping four core documentation artifacts to the specific purpose each serves under both frameworks simultaneously.

Table 5. Unified documentation artifacts serving both SR 11-7 and NIST AI RMF

requirements.

Documentation artifact	SR 11-7 purpose	NIST AI RMF purpose
Model inventory entry	Establishes the model exists, per SR 11-7's inventory requirement, with its materiality assessment and rationale [7][8]	Feeds the Map function's system-context documentation [3]
Validation report	Evidences conceptual soundness, outcomes analysis, and ongoing monitoring under SR 11-7's validation pillar [7][8]	Populates Measure-function quantitative risk assessment [3]
Recalibration log	Evidences effective challenge for each new model version under SR 11-7's validation pillar [7]	Demonstrates Manage-function risk-response action [3]
Governance charter	Establishes board/senior-management oversight under SR 11-7's governance pillar [7][8]	Directly implements the Govern function [9]

The practical value of building documentation this way — as a single artifact serving both frameworks rather than two parallel documentation tracks — is not merely administrative efficiency, though that efficiency is real for a resource-constrained community institution. It also reduces the risk of the two frameworks' documentation drifting apart over time, which could otherwise leave an institution able to satisfy an SR 11-7 examination while having let its NIST AI RMF alignment lapse, or vice versa — a gap that is more likely to develop when two teams maintain separate documentation than when one integrated artifact serves both purposes from the outset.

This unified approach also gives community institutions a practical answer to a resourcing question that recurs throughout this article series: with limited compliance and risk-management staff, should an institution build separate SR 11-7 and AI-governance capabilities, or one integrated function? The documentation-artifact mapping in Table 5 argues for the latter, since the four artifacts it lists are genuinely shared infrastructure rather than four SR 11-7 artifacts plus four separate AI RMF artifacts. A single risk or compliance team, trained to produce and maintain these four artifacts with both frameworks' requirements in view simultaneously, can cover the governance ground that two separately trained, separately resourced teams would otherwise need to cover independently — a meaningful efficiency gain for exactly the resource-constrained community institutions this article series addresses throughout.

LIMITATIONS OF THE EVIDENCE BASE

Several limitations of the evidence assembled here should be stated plainly. SR 11-7 has now stood for well over a decade without a comprehensive revision, and supervisory examination practice on how its principles apply specifically to continuously recalibrated ML systems remains sparse and unevenly documented; the interpretations of proportionality and materiality discussed in Sections 2 and 5 are drawn from industry and legal commentary and this article's own synthesis rather than from a dedicated body of examination precedent addressing adaptive models specifically, and institutions should treat this article's function-mapping and tiering framework as a reasoned starting point rather than an officially endorsed regulatory schema. The NIST AI RMF, while stable since its January 2023 publication, is itself expected to continue evolving with additional guidance, meaning the specific function-mapping in Table 3 may need updating as those materials are published. Finally, this article's institution-tier framework in Table 4 is an illustrative synthesis, not an officially endorsed regulatory schema, and institutions should treat it as a starting point for their own documented tiering rationale rather than a substitute for one.

A further limitation concerns the interaction between SR 11-7 and fair-lending law specifically, an area this article touches on in Sections 2 and 6 but does not treat as its central focus. Model risk management guidance and fair-lending compliance obligations under the Equal Credit Opportunity Act and Regulation B are related but legally distinct requirements, administered in practice by overlapping but not identical teams at most institutions, and this article's synthesis of SR 11-7 and the NIST AI RMF should not be read as a complete fair-lending compliance framework on its own — institutions should treat the governance architecture proposed here as a necessary complement to, not a substitute for, dedicated fair-lending legal review of any credit-decision-influencing model.

EXAMINATION READINESS

Beyond the documentation architecture proposed in Section 7, institutions should prepare specifically for how an examiner is likely to approach a review of an adaptive credit early-warning model under SR 11-7. Drawing on the examiner-focus areas identified in SR 11-7 commentary for community and mid-sized institutions [8], four questions are likely to recur: can the institution produce a complete, current model inventory that includes the early-warning system and any third-party or vendor components it relies on; does the documented materiality assessment for that model reflect an honest assessment of its consequence for borrowers rather than a classification chosen to minimise oversight burden; is the validation evidence for the model's most recent version — not an outdated prior version — available and complete; and does the institution's governance structure demonstrate genuine independence between the team that built or recalibrated the model and the team that validates it,

consistent with SR 11-7's effective-challenge principle [7][8].

Institutions that can answer all four of these questions affirmatively, with documentation to support each answer, are well positioned regardless of how supervisory examination practice continues to develop, since these four questions are drawn directly from principles present in SR 11-7 itself rather than from provisions that might be subject to future interpretive change.

RECOMMENDATIONS

Four recommendations follow for institutions deploying adaptive credit early-warning models under SR 11-7. First, formally document a model materiality assessment under SR 11-7's proportionality principle, treating an adaptive early-warning model as warranting the fullest level of validation given its consequential, fair-lending-relevant use, regardless of the deploying institution's overall asset size. Second, build a single, integrated governance artifact combining SR 11-7 documentation with NIST AI RMF Govern-function accountability structures, rather than maintaining these as separate compliance exercises. Third, budget explicitly for the heightened documentation burden continuous recalibration introduces, treating each recalibration event as requiring its own preserved before/after validation evidence. Fourth, apply the AI RMF's Map function specifically to identify every point at which the early-warning model's output influences a consequential credit decision, since this mapping is the foundation both frameworks rely on to determine appropriate oversight intensity.

CONCLUSION

SR 11-7, now well into its second decade as the foundational U.S. model risk guidance, was written for a generation of largely static, infrequently updated credit and risk models, and does not directly address the continuous recalibration that increasingly characterises adaptive, machine-learning-based credit early-warning systems. For institutions deploying such systems, this gap is an opportunity to build genuinely proportionate — rather than uniformly maximal or uniformly minimal — model risk governance, provided that proportionality is grounded in honest, documented assessment of the model's actual consequence for borrowers rather than used as a rationale for reduced diligence. Combining SR 11-7's supervisory floor with the NIST AI Risk Management Framework's more AI-specific operational detail offers the most defensible path currently available for governing these systems responsibly, pending whatever future AI-specific guidance regulators may eventually issue. For institutions reading this article alongside the others in this series, the practical next step is straightforward: the model-validation, integration-architecture, and fraud-detection articles each describe a specific system component, while this article describes the governance wrapper that makes any one of those components defensible to a regulator. Building the governance architecture proposed here before, or at latest alongside, the

technical deployment work described elsewhere in this series rather than as a retrofit once a system is already in production is the single highest-leverage sequencing decision an institution can make, since retrofitting governance onto an already-deployed adaptive model is materially harder than building it in from the start.

REFERENCES

[1] de Jongh, P. J. (Riaan), Larney, J., Mare, E., van Vuuren, G. W., & Verster, T. (2017). A proposed best practice model validation framework for banks. *South African Journal of Economic and Management Sciences, 20*(1), a1490.

[2] von Thaden, M., & Wehn, C. S. (2020). Model validation and model risk: Reaching the end of the line? *Journal of Banking Regulation, 21*(4), 382–394.

[3] Gourieroux, C., & Monfort, A. (2021). Model risk management: Valuation and governance of pseudo-models. *Econometrics and Statistics, 17*, 1–22.

[4] Kiritz, N., Ravitz, M., & Levonian, M. (2019). Model risk tiering: An exploration of industry practices and principles. *Journal of Risk Model Validation, 13*(4), 1–24.

[5] Farkas, W., Fringuellotti, F., & Tunaru, R. (2020). A cost-benefit analysis of capital requirements adjusted for model risk. *Journal of Corporate Finance, 65*, 101753.

[6] Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics, 57*, 203–216.

[7] Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2020). Explainable AI in fintech risk management. *Frontiers in Artificial Intelligence, 3*, 26.

[8] Ashta, A., & Herrmann, H. (2021). Artificial intelligence and fintech: An overview of opportunities and risks for banking, investments, and microfinance. *Strategic Change, 30*(3), 211–222.

[9] Daníelsson, J., Macrae, R., & Uthemann, A. (2022). Artificial intelligence and systemic risk. *Journal of Banking & Finance, 140*, 106290.

[10] Giudici, P., Centurelli, M., & Turchetta, S. (2024). Artificial intelligence risk measurement. *Expert Systems with Applications, 235*, 121220.

[11] Weber, P., Carl, K. V., & Hinz, O. (2024). Applications of explainable artificial intelligence in finance—A systematic review of finance, information systems, and computer science literature. *Management Review Quarterly, 74*, 867–907.

[12] Ratzan, J., & Rahman, N. (2024). Measuring responsible artificial intelligence (RAI) in banking: A valid and reliable instrument. *AI and Ethics, 4*, 1279–1297.

[13] Hadji Misheva, B., Jaggi, D., Posth, J.-A., Gramespacher, T., & Osterrieder, J. (2021). Audience-dependent explanations for AI-based risk management tools: A survey. *Frontiers in Artificial Intelligence, 4*, 794996.

[14] Lee, J. (2020). Access to finance for artificial intelligence regulation in the financial services industry. *European Business Organization Law Review, 21*, 731–757.

[15] Wirtz, B. W., Weyerer, J. C., & Kehl, I. (2022). Governance of artificial intelligence: A risk and guideline-based integrative framework. *Government Information Quarterly, 39*(4), 101685.

[16] Ahmed, I. E., Mehdi, R., & Mohamed, E. A. (2023). The role of artificial intelligence in developing a banking risk index: An application of adaptive neural network-based fuzzy inference system (ANFIS). *Artificial Intelligence Review, 56*, 13873–13895.

[17] Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: A systematic literature review. *Artificial Intelligence Review, 57*, 216.

[18] Fritz-Morgenthal, S., Hein, B., & Papenbrock, J. (2022). Financial risk management and explainable, trustworthy, responsible AI. *Frontiers in Artificial Intelligence, 5*, 779799.

[19] Lin, E. M. H., Sun, E. W., & Yu, M.-T. (2020). Behavioral data-driven analysis with Bayesian method for risk management of financial services. *International Journal of Production Economics, 228*, 107737.

[20] Rupeika-Apoga, R., & Marano, P. (2023). The risk landscape in the digital transformation of finance and insurance. *Risks, 11*(7), 129.