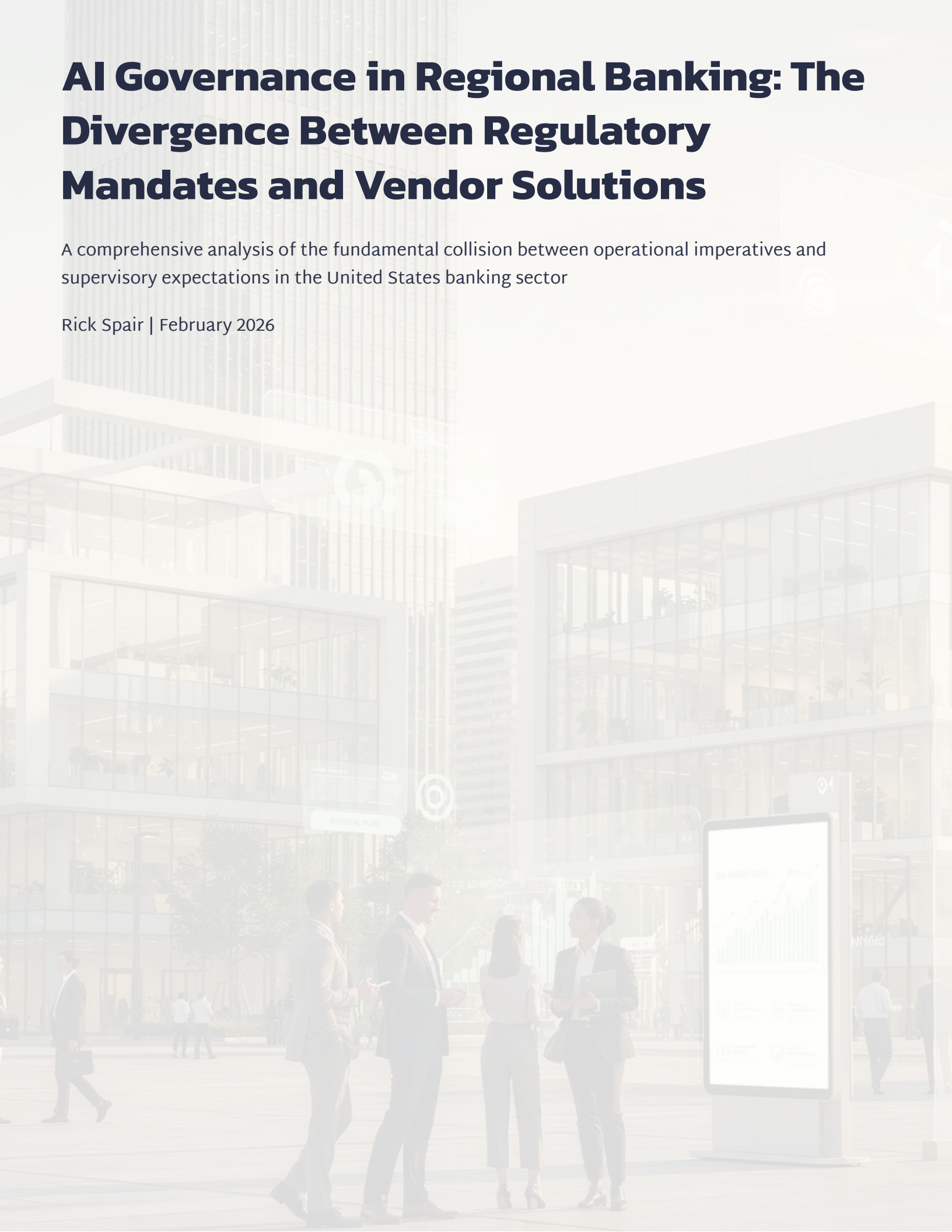
AI Governance in Regional Banking: The Divergence Between Regulatory Mandates and Vendor Solutions

A comprehensive analysis of the fundamental collision between operational imperatives and supervisory expectations in the United States banking sector

Rick Spair | February 2026



Executive Summary

The rapid integration of Artificial Intelligence into the United States banking sector has precipitated a fundamental collision between the operational imperatives of financial institutions and the supervisory expectations of federal regulators. For regional and community banks—institutions with assets between $10 billion and $250 billion—this tension is existential. Unlike Global Systemically Important Banks (G-SIBs), which maintain vast internal cadres of data scientists, model risk auditors, and compliance officers, regional institutions frequently lack the resources to build proprietary AI systems from scratch.

To compete in an increasingly digital economy, these banks are compelled to rely on third-party vendors for advanced analytics in credit underwriting, fraud detection, and anti-money laundering (AML) compliance. This reliance has birthed a profound "Governance Gap"—a distinct and dangerous disconnect between the "automated compliance" solutions marketed by technology vendors and the rigorous demands for "effective challenge" and "conceptual soundness" enforced by the Federal Reserve, the Office of the Comptroller of the Currency (OCC), and the Federal Deposit Insurance Corporation (FDIC).

The vendor landscape in the 2025–2026 cycle is dominated by aggressive marketing claims of "Regulator-Ready" AI, "Fairness-as-a-Service," and "One-Click Documentation." Leading providers such as DataRobot, H2O.ai, Zest AI, and Fairplay AI offer sophisticated platforms designed to abstract away the complexity of machine learning governance. These tools promise to democratize access to advanced modeling, allowing smaller institutions to punch above their weight class.

Critical Reality

Risk management cannot be outsourced. The fundamental regulatory requirement is not merely documentation, but deep institutional understanding and human accountability.

The Regulatory Siege: Understanding the "Want"

To fully grasp the magnitude of the governance gap, one must first rigorously analyze the "Want"—the explicit mandates, implicit expectations, and evolving supervisory philosophies of the U.S. federal banking agencies. Despite the frantic pace of technological innovation in generative and agentic AI, the regulatory framework remains anchored in established principles of safety, soundness, and consumer protection.

The foundational document for AI governance in banking remains the Supervisory Guidance on Model Risk Management, colloquially known as SR 11-7 (Federal Reserve) or OCC 2011-12. Issued in 2011, this guidance was originally drafted to address the failures of financial engineering that precipitated the 2008 financial crisis. Despite its age, regulators continue to cite it as the "Old Testament" of AI governance, applying its tenets to neural networks and large language models with undiminished vigor.

Conceptual Soundness

The model's design, theory, and logic must be supported by empirical evidence and sound judgment

Effective Challenge

Critical analysis by objective, informed parties who can identify limitations and produce changes

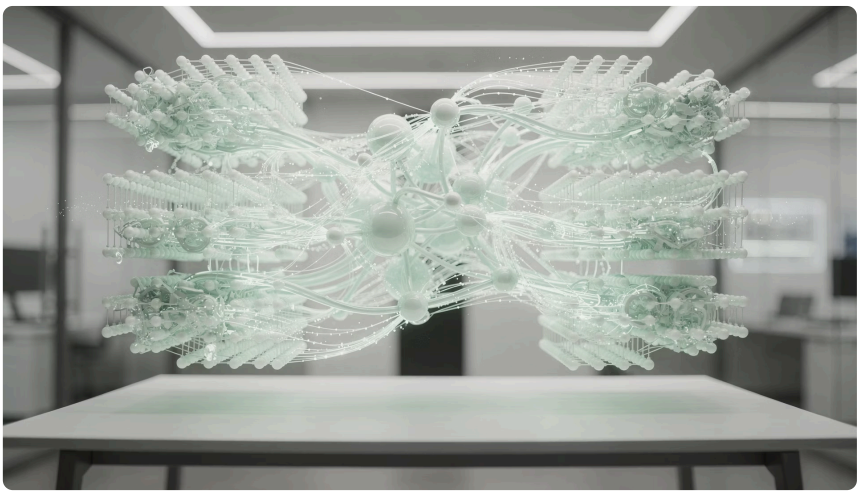
Ongoing Monitoring

Continuous validation that models perform as intended across changing conditions

The Non-Negotiable Standard of Conceptual Soundness

The core pillar of SR 11-7 is "conceptual soundness"—the requirement that a model's design, theory, and logic must be supported by empirical evidence and sound judgment. In the era of linear regression, proving conceptual soundness was a straightforward exercise in statistical transparency. A credit risk model that predicted default based on debt-to-income ratios and FICO scores aligned with economic theory and intuition.

For AI models, particularly "black box" neural networks or ensemble methods used in modern underwriting, proving conceptual soundness is exponentially more difficult. These models often identify non-linear correlations that have no obvious economic rationale. Regulators explicitly reject the notion that complexity excuses opacity. The OCC's Comptroller's Handbook on Model Risk Management emphasizes that if a model cannot be understood by bank personnel, its conceptual soundness is technically unproven.



The AutoML Conflict

The "conceptual soundness" requirement creates an immediate conflict with the "AutoML" (Automated Machine Learning) features marketed by vendors. When a platform automatically selects features and tunes hyperparameters to maximize an accuracy metric, it often does so without regard for economic theory.

> An examiner asking "Why did the model select variable X?" will not accept "Because it improved the Gini coefficient by 0.02" as a sufficient answer. They demand a causal, logical explanation that connects the variable to creditworthiness or fraud risk.

The Mandate for Effective Challenge

Perhaps the most frequently cited deficiency in recent examinations is the lack of "effective challenge." SR 11-7 mandates a critical analysis by objective, informed parties who can identify model limitations and assumptions and produce appropriate changes. This requirement is tripartite, demanding independence, competence, and incentive.

1	2	3
Independence	**Competence**	**Incentive**
The challenge must originate from a party that has no stake in the model's approval or financial success. This precludes the model developers themselves or the business line managers who will profit from the model's deployment.	The challenger must possess the technical expertise to critique the model's architecture. For a regional bank using a sophisticated gradient-boosted tree model for fraud detection, the internal auditor often lacks the specialized data science background to provide a credible challenge.	The challenger must have the authority and the incentive to report issues, even if doing so delays product launches or creates friction with revenue-generating units.

For regional banks relying on third-party AI, "effective challenge" is the primary failure point. If a bank utilizes a vendor's proprietary "black box" model, and relies solely on the vendor's provided validation report, there is no independent challenge. The vendor's report is marketing material, not a risk assessment. Examiners view the uncritical acceptance of vendor validation as a governance failure, often resulting in Matters Requiring Attention (MRAs).

The "Critical Activity" Standard in Third-Party Risk

The regulatory landscape for vendor management was significantly tightened with the release of the Interagency Guidance on Third-Party Relationships: Risk Management, finalized in June 2023. This guidance consolidated fragmented rules into a unified framework that places "critical activities" at the center of the risk map.

An activity is defined as "critical" if it involves significant bank functions (payments, lending, deposit-taking) or if its failure would cause significant harm to the bank or its customers. Under this definition, almost every high-value use case for AI in banking—credit underwriting, fraud detection, AML transaction monitoring, and customer service chatbots—falls under the "critical" umbrella.

1

Diligence Escalator

For critical activities, oversight must be more comprehensive and rigorous than standard vendor questionnaires

2

Fourth-Party Problem

Banks are responsible for subcontractors used by vendors, including cloud providers and foundation model APIs

3

Ongoing Oversight

Continuous monitoring of vendor performance, resilience, and compliance posture

The CFPB's War on Black Boxes

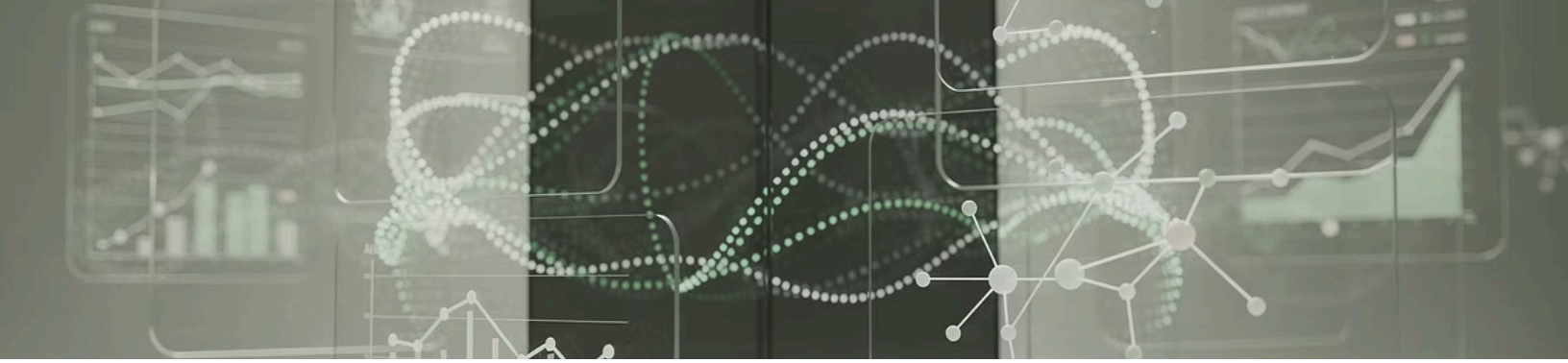
While safety and soundness are the purview of the prudential regulators, the Consumer Financial Protection Bureau (CFPB) has opened a distinct front in the war on opaque AI, focusing on consumer rights and fair lending. In May 2022, the CFPB issued Circular 2022-03, a landmark document that clarified the intersection of the Equal Credit Opportunity Act (ECOA) and complex algorithms.

The End of the Black Box Defense

The circular addressed a specific question: Can a creditor use a complex algorithm if it cannot accurately identify the specific reasons for a credit denial? The answer was a resounding "No." The circular explicitly states that a creditor cannot justify noncompliance with adverse action notice requirements by claiming the technology is "too complicated or opaque to understand."

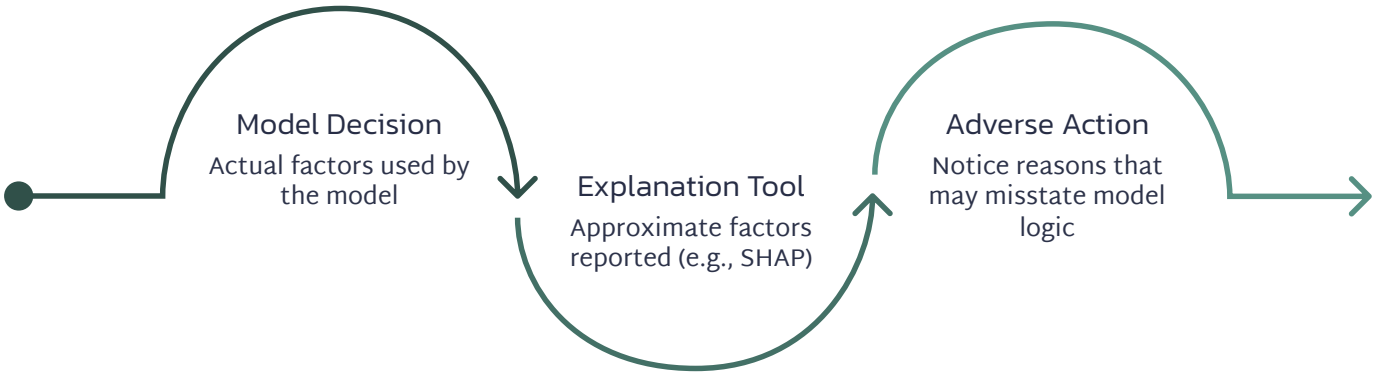
Ignorance of the model's internal mechanics is not a safe harbor; it is a confession of non-compliance. The reasons provided to a consumer for a loan denial must be specific and accurate. Generic checklist codes are insufficient if they do not reflect the actual variables that drove the model's score.

Critical Insight: If a neural network denied a loan primarily because of a complex interaction between "time since last inquiry" and "utilization of revolving credit," the adverse action notice must reflect that specific driver, not a generic proxy.



The Skepticism of Post-Hoc Explainability

The CFPB has also signaled deep skepticism regarding "post-hoc" explainability tools often marketed by vendors, such as SHAP (Shapley Additive Explanations) or LIME. While these tools are mathematically elegant, they are approximations. Research cited in regulatory discussions suggests that these approximations can be unstable—meaning two similar applicants could receive different explanations for the same denial, or the explanation might diverge from the model's actual decision path.



If the explanation tool says the denial was due to "income," but the model actually denied the applicant based on "zip code" (a proxy for race), the bank has committed two violations: fair lending discrimination and providing an inaccurate adverse action notice. Regulators are increasingly testing for this "fidelity gap" during examinations.

The Shift to Outcome Testing and Material Financial Risk

In the 2024-2025 supervisory cycle, there has been a palpable shift in examiner tactics. Moving away from a pure focus on policy documentation ("Does the bank have an AI policy?"), examiners are increasingly focused on Outcome Testing and Material Financial Risk.

Examiners are less interested in reading a vendor's white paper asserting fairness and more interested in seeing the evidence. They expect banks to conduct their own "outcome analysis"—feeding test datasets into the model to verify its behavior across different demographic groups and economic scenarios. The focus is on where AI failure could cause material financial loss, directing scrutiny toward fraud models and credit models.

Evidence Over Assertions

Banks must demonstrate model performance through comprehensive testing, not rely on vendor marketing claims

Financial Impact Focus

Where AI failure could cause material financial loss becomes the primary examination target

Documentation Insufficient

A perfectly documented model that loses money or discriminates will still result in regulatory action

The Vendor Mirage: Analyzing the "Sell"

In direct response to the regulatory complexity described above, a robust industry of "AI Governance" and "RegTech" vendors has emerged. These companies market their solutions as the bridge across the governance gap. However, a granular analysis of their offerings reveals a focus on efficiency, automation, and workflow, which often masks a failure to address the core substantive requirements of regulation.

Leading platforms such as DataRobot, H2O.ai, and ModelOp have positioned themselves as essential infrastructure for banking AI. Their marketing leans heavily on the promise of automating the painful, labor-intensive parts of model risk management. But there exists a critical divergence between vendor capabilities and regulatory expectations.

Automated Documentation	One-click SR 11-7 compliance	Documentation ≠ Challenge
Model Registries	Centralized AI asset visibility	Inventory ≠ Risk Management
AutoML Validation	Automated validation tests	Tests ≠ Conceptual Soundness
Drift Monitoring	Real-time drift alerts	Alerts ≠ Remediation

The Administrative Trap

Vendors are effectively solving the "administrative" problem of compliance—the generation of paper and the tracking of assets. They are not solving the "substantive" problem—the intellectual understanding and effective challenge of risk.

A 500-page automated report that no human at the bank has critically read or understood is a regulatory liability, not an asset.

Fairness-as-a-Service: The Ethics Outsourcing Paradox

Specialized vendors like Fairplay AI and Zest AI have carved out a niche in credit underwriting, offering "Fairness-as-a-Service." These platforms use advanced techniques, such as adversarial debiasing and fairness-aware machine learning, to build models that maintain predictive power while minimizing disparate impact. They offer "regulator-ready" fairness reports and "Fairness Optimizers" that allow banks to toggle between profit maximization and fairness objectives.

However, this creates a dangerous "over-reliance" dynamic. If a bank uses these tools to "optimize" fairness, the bank must still understand what trade-offs were made. Did the model sacrifice 2% of predictive accuracy to reduce the disparate impact ratio by 10%? Did it achieve this by boosting the scores of protected classes using proxy variables?

> Regulators hold the bank accountable for these ethical and legal decisions. A bank cannot say to an examiner, "The software said this was the fair option." The bank must affirmatively justify the trade-off. Outsourcing the mathematics of fairness does not outsource the legal liability.

The 'Regulator-Ready' Marketing Myth

A pervasive theme in vendor marketing is the claim of being 'Regulator-Ready,' 'Audit-Proof,' or 'Compliant with SR 11-7.' This language creates a false sense of security that can lead banks directly into examination failures.

Marketing vs. Reality

"Regulator-ready" is a marketing term, not a legal status. No software is pre-certified by the OCC, Fed, or FDIC. There is no "UL Label" for banking AI.

Contextual Compliance

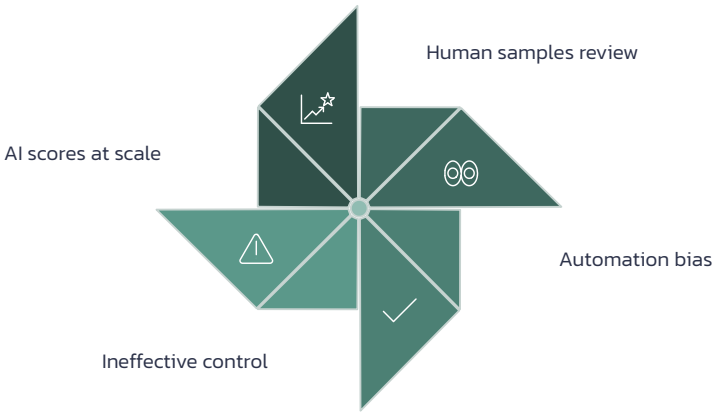
A solution that "passed" an exam at a nimble fintech startup might be deemed insufficient at a conservative regional bank with a lower risk appetite. Examiners tailor their reviews to the specific institution.

False Security

The "Regulator-Ready" label often creates a false sense of security, leading banks to neglect their own due diligence, which is a specific violation of the Third-Party Guidance.

Agentic AI and the Human-in-the-Loop Fallacy

As vendors push "Agentic AI"—autonomous agents capable of executing complex workflows like fraud investigation or customer onboarding—they market "Human-in-the-Loop" (HITL) features as the safety net. However, this presents fundamental challenges that vendors cannot solve through software alone.



The Scale Problem

If an AI agent processes 10,000 transactions per second, real-time human oversight is physically impossible. Vendors implement HITL as a "post-hoc" review or exception handling process, but research and examiner experience show that humans suffer from "automation bias"—they tend to blindly accept the AI's recommendation because "the computer is usually right."

If a human reviewer approves 99.9% of the AI's decisions, examiners do not consider that a valid control. It is a rubber stamp. Vendors sell the interface for human review, but they cannot sell the vigilance required to make it effective.

Critical Friction Points in the Governance Gap

The divergence between regulatory mandates and vendor capabilities creates specific, tangible friction points. These are the areas where regional banks are most likely to fail an examination or face enforcement actions. Understanding these friction points is essential for developing effective risk management strategies.

The Intellectual Property Wall

Direct conflict between banks' regulatory need to validate conceptual soundness and vendors' commercial need to protect trade secrets

Market Power Mismatch

Regional banks lack leverage to negotiate robust contracts with trillion-dollar tech giants

Fourth-Party Opacity

Invisible dependencies on cloud providers and foundation model APIs create unmanaged operational risk

Explainability Inadequacy

Post-hoc tools like SHAP values fail to meet legal requirements for adverse action notices

The Intellectual Property Wall: Code Access vs. Validation

The most acute friction point is the direct conflict between a bank's regulatory need to validate "conceptual soundness" and a vendor's commercial need to protect its Intellectual Property. To fully validate a complex AI model under SR 11-7, a bank theoretically needs access to the source code, training data, and hyperparameter settings. This allows the bank to replicate the model and verify that it works as advertised.

Vendors, especially large fintechs and big tech firms, treat their algorithms as trade secrets. They refuse to share source code or detailed training data, arguing that doing so would expose them to IP theft or competitive replication. The OCC has explicitly stated that a vendor's refusal to provide information does not excuse the bank from its validation duties. If a bank cannot validate a model due to opacity, it is prohibited from using that model for critical activities.

The Regulatory Hardline

A vendor's refusal to provide information does not excuse the bank from its validation duties. Opacity equals prohibition for critical activities.

The Practical Gap

Vendors typically offer "proxy" explanations or white papers describing the model methodology in broad, theoretical terms. Examiners often find these generic documents insufficient for high-risk models, as they do not allow for independent replication or stress testing of the specific model instance used by the bank.

Strategic Responses: The Survival Guide

Faced with these structural challenges, regional banks cannot simply "buy" their way to compliance. They must adopt specific strategic responses that move beyond software to build "governance ecosystems." These strategies represent the practical pathway to managing AI risk in the current regulatory environment.

Compensating Controls

Build proxy models and implement black box testing to validate vendor AI from the outside

Consortium Strategy

Pool resources through alliances like Alloy Labs to gain collective bargaining power and shared intelligence

Governance Culture

Build AI literacy across the organization and establish cross-functional oversight committees

Compensating Controls: Boxing in the Black Box

Proxy modeling stands out as the gold standard for validating vendor AI models when direct access to source code or training data is restricted. This approach involves the bank building its own "challenger model" that aims to replicate the vendor's AI model's output based on observed inputs.

The process entails running both the vendor's AI and the internal proxy model in parallel. By comparing their outputs, the bank can identify and flag any transactions or scenarios where the two models diverge significantly. This divergence serves as a critical alert, prompting further investigation into the vendor model's behavior and performance.

This method not only provides a robust black-box testing framework but also demonstrates to examiners that the bank possesses an independent understanding of the model's functionality and limitations, even without full transparency into its internal workings. It effectively mitigates the risk of using opaque vendor solutions by establishing an internal mechanism for continuous validation and oversight.

The Consortium Strategy: Strength in Numbers

Recognizing that they lack the data scale and negotiating power of Global Systemically Important Banks, regional banks are forming consortia to pool resources and risk intelligence. This represents a fundamental shift in competitive strategy—from isolated competition to collaborative defense against systemic risk.

The Alloy Labs Alliance functions as a "collective bargaining unit" for innovation. Instead of 50 banks each performing a shallow review, the Alliance performs one deep, forensic review. A 50-page technical validation report produced by the Alliance carries significantly more weight with examiners and is more cost-effective than what a single bank could produce.

Fraud data consortia like SardineX solve the "cold start" problem. Regional banks often lack enough fraud data to train robust AI models. By participating in SardineX, a regional bank gains access to a "shared brain" of risk data, training models on millions of outcomes shared by the network. This improves "conceptual soundness" by ensuring the training data is robust and representative. These consortia operate under Section 314(b) of the USA PATRIOT Act, which provides a safe harbor for financial institutions to share information regarding suspected money laundering and terrorist activity.

The Future Regulatory Horizon: 2026 and Beyond

The governance landscape is not static. Two major legislative and policy shifts—the GENIUS Act and the AI Action Plan—are set to redefine the rules of engagement through 2026 and beyond. Regional banks must prepare for a fundamentally transformed regulatory environment.

The GENIUS Act: Audit Culture Spillover

The Guiding and Establishing National Innovation for U.S. Stablecoins Act (GENIUS Act), passed in July 2025, establishes a federal framework for payment stablecoins. While focused on digital assets, it has profound implications for AI governance.

The Act mandates monthly attestations and annual audits for digital asset issuers, creating a regulatory "muscle memory" for strict, frequent auditing of digital tools. Examiners, trained on the GENIUS Act standards, will likely apply this "audit culture" to AI models used for liquidity management and payments. The expectation for "real-time" auditability will grow significantly.

The Act allows non-bank entities (fintechs) to issue stablecoins under federal supervision, significantly increasing competitive pressure on regional banks. To survive, regional banks will be forced to accelerate their adoption of AI for efficiency and customer experience, potentially driving them toward riskier, faster deployments—exactly the behavior examiners will be watching for.

Conclusion: The Path to Augmented Governance

The divide between what regulators want and what vendors sell is not closing; it is widening as AI technology accelerates. Regulators demand human accountability, conceptual understanding, and effective challenge. Vendors sell automation, opacity, and efficiency. For regional banks, the "Governance Gap" is the defining risk of the AI era.

Vendors offer powerful tools that promise to close the technology gap with Global Systemically Important Banks, but often widen the governance gap with regulators. The "Regulator-Ready" sticker on a software package is a mirage—a marketing construct that provides no legal safe harbor and can actively undermine genuine risk management.

Reject the Black Box

Refuse to deploy critical models where the vendor cannot provide sufficient transparency or where the bank cannot build a robust proxy model for validation

Leverage the Collective

Use consortia like Alloy Labs and SardineX to gain the data scale and negotiating leverage needed to manage third-party risk effectively

Invest in Human Capital

Ensure that the "Human-in-the-Loop" is not just a rubber stamp, but a trained risk professional empowered to say "no" to the AI

In the era of Agentic AI, the ultimate regulatory safe harbor is not a software certificate, but a governance culture that can explain—in plain English—why the machine did what it did, and why it was safe to let it do so. This is the path to augmented governance: technology as a tool, not a replacement, for human judgment and institutional accountability.