



RESPONSIBLE AI IN CREDIT PORTFOLIO GROWTH: A RISK-CONTROLLED FRAMEWORK FOR ARTIFICIAL INTELLIGENCE-DRIVEN LENDING IN U.S. BANKS

Yuliia Lukianenko¹

¹*Industry expert in banking and finance, with practical expertise in adapting management and operational processes for U.S. banks and financial institutions*

Abstract

This study develops and validates GUARD (Governance, Underwriting integrity, Accountability, Resilience, Disclosure), a risk-controlled framework for the deployment of artificial intelligence in credit origination and portfolio management by U.S. banking organizations. It examines how lenders can convert AI-driven underwriting into sustained credit portfolio growth while retaining defensible control over fair lending exposure, model risk, and consumer explanation obligations, in a supervisory environment that between April 2025 and June 2026 moved decisively from prescriptive to principles-based expectations.

Methods. A sequential mixed-methods design was employed. Phase one applied structured documentary content analysis to forty-one U.S. federal and state regulatory instruments issued between 2021 and 2026, coded against nine control objectives. Phase two used a design-science approach to construct the framework and a five-level maturity instrument. Phase three tested it through semi-structured interviews with model-risk, credit-risk, fair-lending, and technology executives at U.S. banking organizations (n = 23) and a nine-month maturity assessment at three anonymized mid-size retail banks holding between four and twenty-eight billion dollars in assets.

Scientific novelty. The April 2026 interagency revision of model risk management guidance narrowed the formal supervisory perimeter - rescinding the 2011 framework and its credit-scoring antecedents, introducing a thirty-billion-dollar applicability threshold, and expressly excluding generative and agentic artificial intelligence - precisely as consumer-protection expectations for algorithmic credit decisions were being reaffirmed. Across the three sites the GUARD composite maturity score rose from 2.3 to 3.7 on a five-point scale. Sites recorded a mean 18.4% increase in AI-assisted origination volume and a 22% reduction in time-to-decision over the same nine months; because the design used no randomized or matched control group, these are reported as implementation outcomes observed alongside rising maturity rather than as estimated causal effects of GUARD. The adverse-action explanation defect rate fell from 11.2% to 2.6% and early delinquency on AI-assisted vintages remained indistinguishable from control vintages.

Contributions. The study extends the model risk management literature by theorizing what occurs when the formal supervisory perimeter contracts faster than the technology it governs, a condition termed the governance perimeter gap, and contributes to the responsible artificial intelligence literature by translating a general-purpose risk-management architecture into a sector-specific, auditable instrument for credit decisioning.

Scope of claims. A methodological note on scope applies throughout the study. Three categories of claim recur and are kept distinct: (a) descriptions of the current legal or supervisory status of a specific instrument – binding law, regulation, supervisory guidance, rescinded or stayed guidance, or executive

action – verified against the primary source cited; (b) the authors’ interpretation of what that status implies for institutional obligation, which is argued rather than given; and (c) constructs proposed by the study itself – the governance perimeter gap, the accountability gap, and the GUARD architecture – which are theoretical contributions and not statements of settled law. Table 1a classifies each regulatory source coded in the corpus against this distinction.

The quality of the explanation a declined applicant receives is a distributional matter, not merely a documentation matter. Tying explanation quality to an auditable control connects portfolio growth objectives to the protection of applicants most likely to be assessed by inference rather than by established credit history.

This is the first governance framework for artificial intelligence in lending constructed against the post-April-2026 U.S. supervisory architecture, and the first to treat the contraction of the model risk perimeter as its central design constraint rather than as contextual background.

Keywords

Responsible Artificial Intelligence, Credit Portfolio Growth, Model Risk Management, Fair Lending, Explainable Machine Learning, Adverse Action Notice, Bank Supervision, Algorithmic Underwriting, Governance Maturity, U.S. Banking

Introduction

Artificial intelligence has passed from pilot to production infrastructure in United States lending. Machine-learning underwriting models, cash-flow-based creditworthiness assessment, automated document ingestion, and portfolio-monitoring analytics now sit inside the origination and servicing chains of institutions ranging from money-center banks to community lenders. In its Spring 2026 Semiannual Risk Perspective the Office of the Comptroller of the Currency described banks of all sizes deploying sophisticated models to improve credit underwriting, detect fraud, and personalize customer experience, and characterized a failure to invest in such technology as itself a material risk to long-term viability (OCC, 2026b). By mid-2026 artificial intelligence had become a standing topic in routine examinations, with examiners questioning institutions on model governance, vendor dependency, and operational contingency. Adoption is therefore no longer the strategic variable in lending. Control is.

The commercial logic of adoption is well established. Khandani, Kim, and Lo (2010) demonstrated that machine-learning classifiers materially outperform conventional scorecards in forecasting consumer delinquency; Fuster, Goldsmith-Pinkham, Ramadorai, and Walther (2022) showed that the gains from statistical technology in mortgage lending are real but unevenly distributed across borrower groups; Berg, Burg, Gombović, and Puri (2020) established that digital footprint variables carry predictive content comparable to a bureau score; and Jagtiani and Lemieux (2019) found that alternative data allowed lenders to extend credit to applicants who would have been rated below investment grade under traditional models. Together these findings describe the mechanism through which artificial intelligence generates portfolio growth: not principally by loosening standards, but by extending the frontier of applicants who can be assessed at all.

The control logic is less settled, and the events of 2025 and 2026 have made it less settled still. Executive Order 14281 of April 2025 directed federal agencies to eliminate reliance on disparate-impact liability, and the Office of the Comptroller of the Currency subsequently confirmed through Bulletin 2025-16 that its supervisory process no longer includes examination for it. On 17 April 2026 the three federal banking agencies issued revised interagency guidance on model risk management as SR 26-2, OCC Bulletin 2026-13, and FDIC FIL-16-2026. The revision rescinded the 2011 framework that had governed model risk for fifteen years, together with the 1997 credit-scoring examination guidance and the 2021 interagency statement on anti-money-laundering models. It narrowed the definition of a model, introduced a thirty-billion-dollar asset applicability threshold, and expressly excluded generative and agentic artificial intelligence from its scope, with a request for information on those technologies promised at an unspecified future date (OCC, 2026a). At state level, enforcement of Colorado's pioneering artificial intelligence statute was stayed on 27 April 2026 and the statute repealed and replaced on 14 May 2026 by a narrower disclosure framework effective January 2027.

The same period produced a countervailing signal. On 5 May 2026 the Consumer Financial Protection Bureau issued Circular 2026-03, reaffirming that creditors employing complex algorithms remain fully responsible under the Equal Credit Opportunity Act and Regulation B for furnishing specific and accurate reasons for adverse action. The Bureau's Winter 2025 Supervisory Highlights had already flagged underwriting models incorporating more than a thousand variables, directed institutions to search for less discriminatory alternatives, and stated that federal consumer financial law admits no advanced-technology exception. The resulting configuration is asymmetric: in prudential form supervision has become more permissive and, for institutions below the threshold, largely discretionary, while in consumer-protection substance the obligations attaching to each individual credit decision are unchanged and have been reaffirmed with unusual directness.

This asymmetry produces what this study terms a governance perimeter gap: a condition in which the formal supervisory perimeter contracts, or is drawn to exclude a class of technology, while the legal exposure generated by that technology remains constant. The gap is not a regulatory vacuum but a transfer of the burden of framework design from the supervisor to the institution, and no architecture specific to credit decisioning and mapped to currently operative United States instruments has been articulated in the literature. The central research question is accordingly: how can a United States bank structure the governance of artificial-intelligence-driven lending so that the resulting credit portfolio growth is defensible - evidentially, legally, and prudentially - under a regime simultaneously more permissive in form and more demanding in substance? Three subsidiary questions follow. Which control objectives are supported by operative instruments, and which rest on rescinded or stayed authority? Can a maturity-based framework be granular enough to survive examination yet proportionate enough for mid-size institutions to implement? What measurable effect does implementation have on portfolio growth, decision quality, and consumer explanation performance?

Main Body

The empirical literature has converged on a qualified conclusion: statistical technology improves predictive accuracy, and the resulting gains are distributed unevenly. Fuster and colleagues (2022) characterized machine-learning outcomes in mortgage lending as predictably unequal rather than randomly distributed, since model flexibility allows some borrower groups to be priced more favourably and others less so than under a linear benchmark. Bartlett, Morse, Stanton, and Wallace (2022) found that algorithmic lenders discriminated less than face-to-face lenders in acceptance decisions but continued to extract price differentials from minority borrowers, suggesting that the locus of disparity had shifted from the decision to the price. Gambacorta and colleagues (2024) attributed comparable improvements at a large fintech platform to the volume and granularity of transactional data rather than to algorithmic sophistication as such.

The theoretical implication matters here. Portfolio growth from artificial intelligence derives principally from an expansion of the assessable population, and it is precisely this expansion that generates fair lending exposure, because applicants newly brought within the assessable set are disproportionately those for whom the model must reason by inference rather than by established repayment history. A third and less developed strand concerns systemic effects: the Financial Stability Board (2017) warned that convergence on similar models and shared vendor data could produce correlated behaviour across institutions and amplify procyclicality in credit supply. This concern is largely absent from the bank-level governance literature, which treats model risk as institutional rather than systemic, and it informs the Resilience dimension proposed below.

For fifteen years model risk governance rested on the guidance jointly issued in 2011 as OCC Bulletin 2011-12 and SR 11-7, adopted by the Federal Deposit Insurance Corporation in 2017. It defined model risk as the potential for adverse consequences from decisions based on incorrect or misused model outputs, and organized its mitigation around disciplined development and implementation, independent validation supported by effective challenge, and governance through policies, controls, and a maintained inventory. Bracke, Datta, Jung, and Sen (2019) treated this tripartite structure as the natural home for machine-learning explainability, arguing that interpretability tools belong to validation rather than to consumer disclosure.

Application drifted well beyond original purpose. The Federal Deposit Insurance Corporation observed publicly that the guidance had come to be applied to models posing no material financial risk, burdening smaller institutions and generating criticism directed at documentation rather than at risk. The

April 2026 revision responded directly: the agencies narrowed the definition of a model, adopted the thirty-billion-dollar threshold, clarified proportionality, and stated that the guidance establishes no enforceable standards and that non-compliance will not itself result in supervisory criticism. Most consequentially, they excluded generative and agentic artificial intelligence on the ground that these technologies are evolving too rapidly for stable supervisory treatment. The Federal Reserve's Vice Chair for Supervision framed the revision as a correction of scope rather than a retreat from expectation, observing that banks are relying on existing risk-management frameworks and questioning whether supervisory guidance is fit for the future (Bowman, 2026). What is not in dispute is that the exclusion transfers design responsibility to the institution: an examiner encountering a large-language-model underwriting assistant has no instrument-specific expectation to apply, but retains full safety-and-soundness and consumer-protection authority over the outcome.

The consumer-protection layer has proved markedly more stable than the prudential layer. Regulation B requires that an applicant against whom adverse action is taken receive a statement of the specific principal reasons. The Consumer Financial Protection Bureau addressed complex models in Circular 2022-03, holding that creditors cannot escape the obligation by asserting technological complexity, and in Circular 2023-03, rejecting reliance on standardized reason checklists where those reasons do not accurately describe the basis of the decision. Circular 2026-03 restated the position for machine-learning underwriting platforms, placing responsibility on the lender to understand the algorithm sufficiently to translate its output into specific and accurate grounds for denial. The April 2023 joint statement of four federal agencies had already confirmed that existing civil rights, consumer protection, and competition law applies to automated systems without modification.

The scholarly literature identifies a structural difficulty these instruments do not resolve. Barocas and Selbst (2016) established that discriminatory outcomes arise principally from the specification of the target variable and the composition of training data rather than from deliberate use of protected characteristics, so that excluding protected variables offers little protection; Gillis (2022) developed the point for credit, arguing that regulating model inputs is misdirected because disparate effect is a property of the fitted function. A parallel technical literature addresses explanation. Ribeiro, Singh, and Guestrin (2016) and Lundberg and Lee (2017) provided the post-hoc attribution methods on which most production tooling depends, and Bussmann, Giudici, Marinelli, and Papenbrock (2021) demonstrated their application to credit risk. Rudin (2019) advanced the countervailing argument that for high-stakes decisions inherently interpretable models should be preferred, because an attribution approximates the model rather than stating its reason. The distinction is not academic. Regulation B requires the principal reasons for the decision; a feature attribution describes the marginal contribution of a variable to a score, which is a different object. The gap between them — the accountability gap - is where most examination findings in algorithmic underwriting arise.

General-purpose frameworks proliferated in the preceding period. The Artificial Intelligence Risk Management Framework of the National Institute of Standards and Technology (2023) organizes governance around Govern, Map, Measure, and Manage, and has become the default reference architecture for United States institutions because it is voluntary, sector-neutral, and compatible with existing enterprise risk structures; Mitchell and colleagues (2019) contributed the model card as a documentation artefact; the European Union's Artificial Intelligence Act took the alternative path of binding classification, designating systems evaluating the creditworthiness of natural persons as high-risk. The difficulty with general frameworks is evidentiary rather than conceptual: they specify what should be governed without specifying what must be produced. An examiner reviewing an AI-assisted credit file expects traceable inputs, documented policy rationale, evidence of approval, and demonstrable human judgement; a bank that has adopted a four-function architecture but cannot produce those artefacts for a particular file has satisfied the framework and failed the examination. Sectoral translation - the conversion of general risk-management functions into artefacts constituting admissible evidence in a specific legal context - is the missing step and the design objective here.

Two traditions inform how that translation should be structured. Institutional theory suggests that organizations facing regulatory uncertainty converge on legitimated rather than demonstrably effective forms (DiMaggio & Powell, 1983), that formal compliance structures may become decoupled from operational practice (Scott, 2014), and that the demand for verification reshapes the underlying activity to make it auditable, sometimes at the expense of the substance verified (Power, 1997). These accounts predict a specific failure mode under supervisory ambiguity: institutions adopt the visible artefacts of control while operative decisions remain unexamined. The resource-based view explains why an institution

would invest beyond compliance. Barney (1991) locates advantage in resources that are valuable, rare, imperfectly imitable, and non-substitutable, and Teece, Pisano, and Shuen (1997) extend this to the capacity to reconfigure competencies in changing environments. A governance architecture capable of certifying an AI-driven credit decision as defensible is such a capability: costly to build, difficult to imitate because embedded in data infrastructure rather than documentation, and directly enabling of growth, since the binding constraint on deploying a profitable model is frequently not its predictive performance but the institution's capacity to defend its use.

Propositions

P1: a contraction of the formal supervisory perimeter that is not matched by a corresponding contraction in consumer-protection or civil-liability exposure produces a governance perimeter gap, operationalized as the divergence between an instrument's coded status (operative, rescinded, or stayed) and the continued enforceability of the obligations it addressed. P2: under a governance perimeter gap, institutions that rely on supervisory attention to prioritize controls – the decoupling mechanism described by DiMaggio and Powell (1983) and Scott (2014) – will underinvest in file-level, individually enforceable controls relative to portfolio-level prudential controls, because the latter retain greater supervisory salience even where the former carries the greater expected liability. P3: an accountability gap – the divergence between a post-hoc feature attribution and a legally sufficient statement of principal reasons under Regulation B – is reduced when the attribution and the communicated reason are generated by the same process step rather than reconstructed afterward, independent of the sophistication of the attribution method itself. P4: governance maturity constrains deployable capacity rather than model performance, so that increases in AI-assisted origination volume following governance investment are expected to reflect release of a pre-existing deployment backlog rather than a causal treatment effect on demand or approval propensity, and to decay as the backlog clears. P5: governance investment functions as a dynamic capability (Teece, Pisano, & Shuen, 1997) only where the resulting evidence is a by-product of the operational process rather than a documentation exercise layered onto it (Power, 1997); domains structurally coupled to the credit-decision workflow, such as Accountability, should therefore show a different maturation path than domains that are not, such as Resilience. These propositions are evaluated, not confirmed, by the single-country, three-site evidence reported in sections 4 and 5; they are offered as a research agenda as much as a finding, and each is revisited in the Discussion.

The study employs a sequential exploratory mixed-methods design (Creswell & Plano Clark, 2018) in three phases: documentary content analysis to establish the regulatory basis, design-science construction of the artefact (Hevner, March, Park, & Ram, 2004), and evaluation through practitioner validation and field implementation. The sequence is deliberate. Because the central problem is instability of the regulatory perimeter, the framework had to be derived from the corpus rather than imposed on it, so that each control could be traced to an operative instrument and each rescinded authority explicitly identified.

A corpus of forty-one instruments was assembled, comprising federal supervisory guidance, agency circulars and bulletins, interagency statements, executive orders, supervisory publications, and state legislation issued between March 2021 and June 2026. The starting boundary was the March 2021 interagency request for information on financial institutions' use of artificial intelligence, the first coordinated federal engagement with the subject. Instruments were included where they addressed model governance, algorithmic decision-making, fair lending in automated systems, third-party technology risk, or consumer disclosure in credit decisions, and each was coded on current status (operative, rescinded, stayed, or pending), addressee class, and the control objectives it supports. Nine objectives were derived inductively across two coding passes separated by six weeks: inventory and tiering, data provenance, development discipline, independent validation, human override, decision explanation, outcome monitoring, third-party oversight, and consumer disclosure.

The nine objectives clustered into five domains. Governance addresses ownership, inventory, risk tiering, and board reporting. Underwriting integrity addresses data provenance, development discipline, independent validation, and effective challenge. Accountability addresses translation of model output into legally sufficient reasons, the operation and evidencing of human override, and the file-level audit trail. Resilience addresses drift monitoring, portfolio concentration, stress testing of model-dependent segments, and the capacity to suspend a model in production. Disclosure addresses consumer notice, vendor transparency, and supervisory communication.

Each domain is assessed on a five-point maturity scale. Level one denotes ad hoc practice with no documented control; level two, documented policy without systematic evidence of operation; level three, consistent operation with evidence produced on request; level four, evidence generated automatically as a by-product of the model lifecycle with independent challenge functioning; level five, continuous measurement with demonstrated remediation cycles integrated into credit strategy. The distinction between levels three and four is the pivot of the instrument and follows from the concern identified in section 2.4: an institution that assembles evidence on request has a documentation function, whereas an institution whose evidence is generated by the process itself has a control.

Scale construction and scoring reliability

Each of the nine control objectives was scored individually against the same five-point definitions, and a domain score was calculated as the unweighted mean of its constituent objectives; objectives were not differentially weighted, because the interview sample did not converge on a defensible weighting scheme and an unweighted mean was judged more transparent and easier to audit than an undisclosed one. Each level within each objective was anchored to observable evidence requirements developed iteratively with the twenty-three interview respondents – for example, at Accountability level four, evidence required a system-generated log linking the file-level attribution record to the communicated reason code at the moment of decision, not a policy stating that such linkage should occur. At each site and assessment wave, two raters – the author and one senior risk officer at the site who had no role in the framework's design – independently scored each objective against the evidence produced and then reconciled disagreement through discussion; unreconciled agreement across the 108 scoring instances (nine objectives × three sites × four waves) was 78%. This procedure improves transparency over an unaudited single-rater score but falls short of a validated psychometric instrument: formal inter-rater reliability statistics computed prior to reconciliation, and replication by raters with no relationship to the framework's development, were not obtained here and are identified in section 5.4 as a priority for independent validation.

GUARD Domain	Principal Control Objectives	Operative Regulatory Anchors (2026)	Control Intensity
Governance	Inventory, risk tiering, ownership, board reporting	SR 26-2 / OCC 2026-13 / FDIC FIL-16-2026; NIST AI RMF 1.0	Moderate
Underwriting integrity	Data provenance, development discipline, independent validation, effective challenge	SR 26-2 (below threshold: by analogy); ECOA / Reg. B; CFPB Suppl. Highlights, Winter 2025	High
Accountability	Specific reasons for adverse action, human override, file-level audit trail	CFPB Circulars 2022-03, 2023-03, 2026-03; 12 C.F.R. § 1002.9	Very high
Resilience	Drift monitoring, portfolio concentration, stress testing, model suspension	Safety-and-soundness authority; OCC Semiannual Risk Perspective (2026)	Moderate
Disclosure	Consumer notice, vendor transparency, supervisory communication	Interagency Third-Party Risk Guidance (2023); CO SB 26-189 (eff. 2027)	High

Table 1. GUARD Domains, Control Objectives, and Operative Regulatory Anchors

Regulatory source classification

Table 1a implements the classification promised above: each regulatory source used to construct Table 1 and the surrounding argument is coded by legal category and current status, so that the claim of a governance perimeter gap rests on a visible distinction between what is binding, what is guidance, and what is proposed.

Instrument	Legal category	Current status (as of June 2026)	Function in the governance-perimeter argument
ECOA / Regulation B (12 C.F.R. Part 1002)	Binding statute / regulation	Operative, unaffected by 2025-2026 changes	Source of the file-level obligation that the perimeter gap argument holds constant
SR 11-7 / OCC Bulletin 2011-12 (2011 model risk framework)	Supervisory guidance	Rescinded, April 2026	Removal of the tripartite architecture that anchored Underwriting integrity and Governance
OCC Bulletin 1997-24 (credit-scoring examination guidance)	Supervisory guidance	Rescinded, April 2026	Removal of the only instrument specific to credit-scoring model examination
SR 26-2 / OCC 2026-13 / FDIC FIL-16-2026 (2026 revision)	Supervisory guidance (self-described as non-binding)	Operative; states non-compliance is not itself grounds for criticism	Defines the narrowed formal perimeter (threshold, model definition, generative/agentive exclusion)
CFPB Circulars 2022-03, 2023-03, 2026-03	Interpretive / agency statement	Operative	Reaffirms the Accountability obligation independent of the prudential perimeter
Executive Order 14281 (April 2025)	Executive action	Operative; directs agency practice, does not amend statute	Basis for OCC Bulletin 2025-16; does not itself alter ECOA or Reg. B
OCC Bulletin 2025-16	Supervisory / examination-practice guidance	Operative	Removes disparate-impact review from OCC examination scope only
Interagency Guidance on Third-Party Relationships (2023)	Supervisory guidance	Operative	Anchors the Disclosure domain's vendor-transparency objective
NIST AI RMF 1.0 (2023)	Voluntary framework	Not legally binding	Reference architecture only; not treated as an obligation in Table 1
Colorado SB 26-189	State statute	Enacted 2026; effective January 2027; narrower than the repealed 2024 statute	Illustrates state-level contraction of the AI-specific perimeter, coded separately from federal instruments
2023 interagency/DOJ/FTC/EEOC joint statement	Interpretive / agency statement	Operative	Confirms existing civil-rights and consumer law applies to automated systems without modification

Table 1a. Classification of Regulatory Sources by Legal Category and Status

The draft framework was tested through twenty-three semi-structured interviews conducted between October 2025 and April 2026 with model-risk, credit-risk, fair-lending, and technology executives

across three asset tiers: above one hundred billion dollars ($n = 7$), between ten and one hundred billion ($n = 9$), and below ten billion ($n = 7$).

Recruitment, protocol, and coding

Recruitment proceeded through purposive, network-based sampling: respondents were identified through professional banking-governance networks and industry-conference contacts, screened for a current model-risk, credit-risk, fair-lending, or technology leadership role with direct responsibility for AI governance, and invited by email; twenty-three of thirty-one invited executives (74%) agreed to participate, and no incentive was offered. This recruitment approach is subject to a selection concern addressed in the limitations: participating institutions and individuals were, on average, more likely to already hold governance capability in some form, and the sample should not be read as representative of the sector. Interviews lasted 45 to 60 minutes, followed a common protocol covering the respondent's current governance architecture, perceived gaps under the 2025-2026 regulatory changes, and reaction to successive drafts of the GUARD domain structure, and were recorded and transcribed with consent. Transcripts were coded thematically by the author following Braun and Clarke (2006); a second coder, a doctoral researcher in financial regulation with no role in the framework's design, independently coded a 30% random subsample, yielding 81% agreement on theme assignment at the level of the coded segment, with disagreements resolved by discussion. Coding continued until no new first-order themes emerged across the final three interviews, an indicative rather than a formally tested saturation criterion given the sample size. The twenty-three interview respondents and the three implementation sites are analytically distinct but not independent – two of the three sites each contributed one interview respondent to the twenty-three – and the reader should note that the site-level maturity and outcome data reported in section 4 are derived from document review and control walkthroughs, not from interview transcripts; interview data are used only for framework refinement and for the thematic interpretation offered in the Discussion.

Two structural revisions followed. Human override was relocated from Governance to Accountability, on the reasoning that an override is evidentially meaningful only when it attaches to an individual credit file; model suspension capability was elevated to a scored element of Resilience, following the observation that examiners had begun asking about kill-switch arrangements in routine examinations.

Three mid-size retail banks were recruited as implementation sites through purposive sampling, on criteria of consumer or small-business lending as a primary line, at least one machine-learning model in production within the credit decision chain, total assets between one and thirty billion dollars, and no prior adoption of a formal artificial intelligence governance framework. Site A holds approximately four billion dollars and operates a single model supporting unsecured consumer lending. Site B holds approximately eleven billion and operates four models, two vendor-supplied. Site C holds approximately twenty-eight billion, sits immediately below the threshold, and operates seven models including a vendor-hosted document-processing system with generative components.

Baseline assessment was conducted in the third quarter of 2025 through document review, control walkthroughs, and interviews with domain owners, with reassessment at three-month intervals over nine months. Quantitative data comprised origination volume and approval rates for portfolios subject to AI-assisted decisioning; mean time-to-decision; delinquency by origination vintage, with AI-assisted vintages compared against contemporaneous vintages decisioned without model assistance; adverse-action explanation defect rates, defined as the proportion of sampled notices in which the stated principal reasons could not be reconciled to the model's file-level attribution record; validation cycle time; and documented override frequency. Qualitative data comprised implementation interviews at months three, six, and nine, analysed thematically following Braun and Clarke (2006). Two constraints require statement: portfolio data were site-supplied and not independently audited, and the nine-month window without a randomized control supports inference about association rather than causation.

Validating the explanation-defect measure

The explanation-defect determination requires separate comment, since a technical inconsistency between a stated reason and a model's attribution record is not, by itself, equivalent to legal insufficiency under Regulation B; the two were kept analytically distinct. A notice was coded as technically defective where the top-ranked reason codes communicated to the applicant did not appear among the model's highest-

magnitude attributions for that file, a reconciliation performed by a compliance analyst at each site who was blinded to the site's current maturity score at the time of coding but not blinded to institution, since cross-site blinding was not operationally feasible. A narrower category, legal review, was applied where a fair-lending or compliance officer additionally assessed whether the communicated reasons, whatever their technical provenance, were specific enough to satisfy Regulation B's principal-reasons standard; legal review covered the full sample at Site B and a 25% sample at the other two sites, so the explanation-defect rate reported in Table 3 is the technical-reconciliation rate and should not be read as a rate of adjudicated legal violations. In the Site B full sample, where both measures were available, technical defects outnumbered legally deficient notices by roughly three to one – consistent with the concern that reconciliation failure and legal insufficiency are related but not identical constructs – and this divergence is reported as a finding in section 4.5 rather than resolved by assumption.

The documentary analysis produced a finding not visible from any single instrument. Of the forty-one instruments coded, eleven were rescinded, superseded, stayed, or replaced during the corpus period, and of these, eight had previously been treated as primary authority for artificial intelligence governance in lending. The rescission of OCC Bulletin 2011-12 removed the tripartite architecture; the rescission of Bulletin 1997-24 removed the only instrument addressing credit scoring model examination specifically; the stay and repeal of Colorado's statute removed the sole comprehensive state framework covering credit as a consequential decision; and Executive Order 14281 with Bulletin 2025-16 removed disparate impact from the supervisory examination process. The instruments that remained operative are consumer-protection instruments applying to individual decisions, not prudential instruments applying to model portfolios.

This produces a distinctive risk profile. The surviving obligations attach at the file level and are enforceable through private litigation as well as through supervision. Regulation B liability does not depend on examination priorities, nor does the Fair Housing Act, nor do state unfair and deceptive practices statutes; removing disparate impact from the Comptroller's examination process narrows one channel of scrutiny without altering the underlying statutory text, and does not bind private plaintiffs or state attorneys general. Several respondents at larger institutions articulated this clearly: reduced examination pressure had, in their assessment, increased rather than decreased the importance of maintaining testing evidence, because the evidence's primary consumer had shifted from the examiner to opposing counsel.

The exclusion of generative and agentic systems from the model risk perimeter emerged as the single most consequential feature for the sites. All three operated at least one system falling outside the revised definition of a model while nonetheless influencing credit outcomes - document extraction feeding underwriting inputs at Site B, a memo-drafting assistant at Site C, and a customer-facing pre-qualification interface at Site A. Under the pre-2026 framework these had been treated, with varying conviction, as models; under the revised framework they are not, yet their outputs still reach the credit file. Site C's chief risk officer described the position as being told that a control does not apply while remaining fully responsible for the outcome it was designed to prevent. GUARD addresses this by tiering on credit impact rather than technical classification, so that a system influencing a credit decision is scoped in regardless of whether it satisfies the definition of a model.

Composite maturity across the three sites rose from a baseline mean of 2.3 to 3.7 over nine months, improving in every domain at every site.

GUARD Domain	Site A (\$4bn)	Site B (\$11bn)	Site C (\$28bn)
Governance	2.4 → 3.6	2.6 → 3.8	2.8 → 4.2
Underwriting integrity	2.0 → 3.8	2.2 → 4.0	2.6 → 4.4
Accountability	1.8 → 3.4	2.0 → 3.6	2.2 → 3.8
Resilience	2.2 → 3.2	2.4 → 3.6	2.6 → 4.0
Disclosure	2.0 → 3.2	2.2 → 3.4	2.4 → 3.6
Composite	2.1 → 3.4	2.3 → 3.7	2.5 → 4.0

Table 2. GUARD Maturity Scores by Implementation Site (Baseline → 9 Months)

Underwriting integrity recorded the largest absolute gain at every site, rising by a mean of 1.8 points; it is the most technically specified domain and the one for which the rescinded 2011 framework had produced fifteen years of accumulated practice, so that sites were formalizing controls they already partially

operated. Accountability recorded the lowest terminal scores despite substantial improvement, reflecting the difficulty of moving from level three to level four where the required artefact - a file-level record reconciling the model's attribution to the reasons actually communicated to the applicant - must be produced at the moment of decision and cannot be reconstructed. Disclosure improved least, attributed in interviews almost entirely to dependencies on third parties whose contractual terms did not require the transparency the framework demanded. Site C achieved the highest terminal composite, consistent with the resource explanation, though its leadership also anticipated crossing the threshold within its planning horizon and treated the framework as preparation for a regime it expected to enter rather than as a response to one it faced.

Indicator	Baseline	9 Months	Change
AI-assisted origination volume (indexed)	100.0	118.4	+18.4%
Approval rate, AI-assisted segments (%)	31.2	34.6	+3.4 pp
Mean time-to-decision (hours)	41.5	32.4	-21.9%
30+ DPD, AI-assisted vintages (%)	2.41	2.38	-0.03 pp
30+ DPD, control vintages (%)	2.36	2.44	+0.08 pp
Adverse-action explanation defect rate (%)	11.2	2.6	-76.8%
Mean validation cycle time (weeks)	14.6	9.2	-37.0%
Documented override rate (% of overrides)	38.0	94.5	+56.5 pp

Table 3. Aggregate Portfolio and Control Indicators (Baseline vs. 9 Months)

Interpreting Table 3

A design point governs the interpretation of every indicator in Table 3. The implementation was a single-arm, three-site field study without a randomized or matched comparison institution and without a statistical model controlling for portfolio composition, macroeconomic conditions, lending demand, staffing, or the concurrent regulatory developments described in section 1. The changes reported below are therefore implementation outcomes observed over the same nine months in which GUARD maturity rose – a temporal association – and not causal treatment effects of GUARD, however consistent the direction of change across sites. Confidence intervals and significance tests are not reported because a site-level *n* of three is too small to support them meaningfully; site-level variation is instead described qualitatively where it is informative. Where the discussion below proposes a mechanism, that mechanism is the interpretation most consistent with the qualitative interview evidence, not a statistically identified causal pathway.

The origination result requires careful interpretation, since it would be convenient and incorrect to attribute an eighteen per cent volume increase to a governance framework. The mechanism reported consistently across sites was narrower and more credible: governance maturity removed a constraint on deployment rather than improving model performance. At baseline all three sites had models that had completed development but were held out of production, or confined to restricted segments, because the institution could not satisfy itself that deployment was defensible; Site B had two such models, one for eleven months. As Underwriting integrity and Accountability maturity rose, these models entered or expanded production. The growth is therefore a release of accumulated deployment backlog and should be expected to decay once the backlog clears - a point of practical importance, since a bank projecting the observed rate forward will overstate the return on further governance investment.

The delinquency comparison is the most important safety finding and the least dramatic. Early delinquency on AI-assisted vintages was flat while control vintages deteriorated slightly, a difference well within the range attributable to portfolio composition and macroeconomic drift over nine months. The appropriate conclusion is not that governance improved credit quality but that expansion of AI-assisted origination did not degrade it - the relevant question, since the principal prudential concern about model-driven growth is that volume gains are purchased with unrecognized risk surfacing in later vintages. The observation window is too short to exclude that possibility.

The explanation defect rate is the clearest result. A baseline rate above eleven per cent means that in more than one adverse action notice in nine, the reasons communicated to the applicant could not be

reconciled to the model's own record of what drove the decision. Since Regulation B requires the specific principal reasons, a defect of this kind is not a documentation lapse but a probable violation replicated across every affected file. The reduction to 2.6% was achieved through a single architectural change adopted at all three sites: the attribution record and the reason codes were generated in the same process step, from the same object, at the moment of decision, rather than the reason code being selected downstream from a standard list. This is precisely the transition from level three to level four maturity, and it is the strongest evidence for the design principle that evidence must be a by-product of the process rather than a reconstruction of it.

The Accountability findings locate the accountability gap in operational terms. All three sites deployed post-hoc attribution methods and initially treated their output as satisfying the adverse action requirement. File review showed why this fails. An attribution ranks the marginal contribution of features to a score relative to a reference population; the statutory requirement is a statement of the principal reasons the applicant was declined. Where a decision turns on interactions among correlated variables, the highest-ranked attributions may be individually uninformative and may not identify the condition the applicant would need to change. Site B's remediation is instructive: rather than adopting a more sophisticated attribution method, it constrained the model's feature space in the segments where explanation quality was weakest, accepting a small measured reduction in discriminatory power in exchange for reasons that could be stated. In the affected segment, Site B reported a decline in model discriminatory power (Gini coefficient) from 0.61 to 0.57, a 1.1-percentage-point reduction in the segment approval rate, and a fall in the segment's explanation-defect rate from 9.8% to 1.4%, measured over the two quarters following the restriction. These site-reported figures were not independently audited and describe a single institution's single remediation; they illustrate rather than establish the interpretability-accuracy trade-off discussed in the literature (Rudin, 2019), and are reported here as a descriptive case rather than as generalizable evidence of the trade-off's magnitude. This is Rudin's (2019) argument arrived at empirically and under commercial pressure, and it suggests the interpretability-accuracy trade-off in credit is more favourable than the technical literature implies once the cost of unexplainable decisions is priced.

Resilience produced the weakest engagement, for structural rather than technical reasons. Every element except operational contingency addresses risk materializing over a horizon longer than the sites' planning cycles: drift monitoring produces value when a model degrades, concentration analysis when correlated exposures move together, and neither yields observable benefit in nine benign months. Respondents acknowledged the domain's importance and deprioritized it in practice. The Financial Stability Board's (2017) concern about correlated model behaviour across institutions received no substantive attention at any site, which is unsurprising, since no individual institution has either the data or the incentive to assess it - a genuine limitation of institution-level frameworks, including this one.

Vendor dependency emerged as the binding constraint on Disclosure maturity and, at two sites, on Underwriting integrity. Sites B and C together operated nine models, five vendor-supplied or vendor-hosted, and in four of the five the contractual arrangement did not entitle the institution to the documentation the framework required - training data composition, validation methodology, and the record needed to reconcile a decision to its reasons. The Interagency Guidance on Third-Party Relationships (2023) is unambiguous that a bank cannot outsource responsibility for compliance, and the revised model risk guidance does not alter that position, so a mid-size institution's governance maturity is bounded above by its procurement leverage. Site C renegotiated two contracts at renewal to include audit and documentation rights and reported that the vendor accommodated the request without material price effect once it was raised as a condition of renewal; Site B, lacking comparable leverage, restricted one vendor model to advisory use and documented that restriction as a compensating control. Both responses are defensible; they are not equally available. This bears on the proportionality logic of the April 2026 revision, which tailors expectations to size on the assumption that smaller institutions have simpler model estates. For vendor-dependent institutions the assumption inverts: a small bank running vendor artificial intelligence may face a more opaque estate than a large bank building its own, while possessing less capacity to compel transparency.

A conceptual model of the argument

Table 4 makes explicit the conceptual chain that sections 2.5, 4, and 5 have argued piece by piece, so that it does not need restating in full at each step below. Only the third and fourth links are directly measured in this study; the first two links are argued from the documentary and interview evidence rather than statistically identified, consistent with the caution set out in section 4.1.

Link	Mechanism	How it is evidenced here	Status of the claim
1. Regulatory change → governance perimeter gap	Rescission/exclusion narrows the formal supervisory perimeter faster than file-level consumer-protection obligations contract	Documentary content analysis of 41 instruments (Table 1a)	Argued from primary sources, not statistically tested
2. Perimeter gap → GUARD control demand	Institutions must construct evidence for domains no longer specified externally (P1, P2)	Interview evidence (n = 23); design-science construction of GUARD	Theoretical proposition, illustrated by interview themes
3. GUARD controls → governance maturity	Implementing the nine control objectives raises the five-level maturity score	Directly measured: Table 2 (2.3 → 3.7 composite)	Directly measured, single-rater-pair scoring (78% pre-reconciliation agreement)
4. Maturity → deployment capability → portfolio outcomes	Rising maturity releases models held out of production (backlog release, P4), which appears as origination volume, time-to-decision, and explanation-defect change	Directly measured: Table 3, interpreted via interviews	Association, not a causal treatment effect (section 4.1)

Table 4. Conceptual Model Linking Regulatory Change, GUARD Controls, Governance Maturity, and Portfolio Outcomes

The prevailing framing treats governance as a cost imposed on growth. The evidence here supports a different relationship under a contracted perimeter. Where regulatory expectation is prescriptive and stable, governance is largely a compliance cost, because the required controls are specified externally. Where the perimeter is contracted or expressly excludes the relevant technology, the institution must construct and defend its own architecture, and the quality of that architecture becomes the operative constraint on how much of its model estate it can safely deploy. In that condition governance is not a tax on growth but its enabling condition - the mechanism visible in the deployment backlog release. This specifies the content of the dynamic capability in question: not the ability to build models, increasingly commoditized through vendor platforms, nor the ability to write policy, imitable at low cost, but the ability to generate, as an automatic by-product of model operation, the evidence required to defend a decision to an examiner, a court, or a counterparty.

For model risk management scholarship, the study names a condition the existing literature, developed entirely within a period of expanding supervisory scope, has not addressed. The prevailing assumption is that governance quality is driven by supervisory pressure; the evidence here suggests that under a contracted perimeter the driver shifts to litigation and counterparty exposure, altering which controls institutions prioritize - toward file-level, individually enforceable obligations and away from portfolio-level prudential ones. The neglect of Resilience across all three sites is the observable signature of that shift. For institutional theory, the results support the decoupling prediction while identifying a boundary condition: decoupling is unstable where the artefact of compliance is also an operational necessity. Reason codes must be generated for every declined application regardless of maturity; the only question is whether they come from the model's own record or from a list. Where a compliance artefact sits on the critical path of the business process, the framework's demands cannot be satisfied ceremonially.

For practitioners, the mapping in Table 1 distinguishes controls supported by operative authority from those resting on rescinded or stayed instruments, and the ranking is counterintuitive: Accountability, the domain with the least prudential guidance attaching to it, carries the highest control intensity, because its obligations are individually enforceable, apply to every credit decision without threshold, and were expressly reaffirmed in May 2026. An institution allocating governance resource in proportion to supervisory attention will systematically underinvest in exactly the domain where its exposure is greatest. For institutions below the threshold, a maturity instrument in which each level is defined by the evidence it produces allows a deliberate proportionality judgement to be documented - the position the revised guidance in fact invites. For vendor-dependent institutions, documentation and audit rights are obtainable at contract renewal at little or no cost and frequently unobtainable afterwards at any cost, so governance requirements belong in the procurement cycle rather than the risk-review cycle.

The sample comprises three institutions in one national market, all mid-size, all below the supervisory threshold, and all self-selected on a pre-existing willingness to invest in governance; large institutions with mature model risk functions and small lenders with minimal model use are both outside it. Nine months is sufficient to observe control maturation and short-term portfolio effects but insufficient to observe the credit performance of the vintages originated during the study, which is the outcome that would most directly test whether governance-enabled growth was prudent. The absence of a control group limits causal inference, concurrent regulatory change means site behaviour cannot be attributed to the framework alone, and performance data were site-supplied. The author's dual role as framework developer and researcher introduces a risk of confirmation bias, partially mitigated by scoring against defined evidentiary criteria at fixed intervals but not eliminated. Finally, the regulatory environment remains in motion: the interagency request for information has been announced but not issued, Colorado's replacement statute takes effect in January 2027 with rulemaking incomplete, and the litigation posture of federal authorities toward state artificial intelligence legislation is unresolved. Any of these could alter the mapping in Table 1 without altering the framework's structure.

A note on what the evidence supports

The findings summarized below should be read as evidence from a single-country, three-site, nine-month exploratory field study, conducted in part by the framework's own developer and without a control group, sufficient to establish plausibility and candidate mechanisms but not sufficient to establish a generalizable causal effect. The propositions set out in section 2.5 are offered for independent testing rather than as conclusions this study has itself confirmed.

This study asked how a United States bank can govern artificial-intelligence-driven lending so that the credit portfolio growth these systems generate is defensible under a supervisory regime simultaneously more permissive in form and more demanding in substance. The answer developed here is that defensibility must be engineered into the model lifecycle rather than assembled around it, and that the operative test is whether the evidence required to justify a credit decision is produced by the process that makes the decision or reconstructed afterwards for review.

The regulatory reset of 2025 and 2026 makes this urgent rather than academic. The rescission of a fifteen-year-old model risk framework, the introduction of an asset threshold, the exclusion of generative and agentic systems, and the removal of disparate impact from supervisory examination together narrowed the formal perimeter substantially, and the obligations that survived are those attaching to individual credit decisions and enforceable outside the supervisory channel. The consequence is not less governance but differently targeted governance, weighted toward file-level accountability and away from the portfolio-level prudential controls the rescinded framework emphasized. GUARD operationalizes this reweighting across five domains and a five-level maturity scale in which each level is defined by the evidence it produces. In a nine-month implementation at three mid-size retail banks, composite maturity rose from 2.3 to 3.7, AI-assisted origination volume rose 18.4% alongside release of accumulated deployment backlog – an association observed over the implementation window, not an estimated causal effect – early delinquency on AI-assisted vintages remained flat against control vintages, and the adverse-action explanation defect rate fell by more than three quarters.

Four directions for further research follow. Longitudinal work tracking the credit performance of vintages originated under elevated governance maturity over three to five years would test the central prudential question this study could not answer. Extension to institutions above the supervisory threshold would establish whether the domain weightings hold when prudential and consumer-protection obligations

are both fully operative. The vendor-dependency finding suggests a research agenda on procurement as a governance mechanism. Most consequentially, the correlated-model-behaviour concern raised by the Financial Stability Board (2017) and neglected by every site in this study cannot be addressed at institutional level and requires supervisory data that does not presently exist in the public domain. The interagency request for information, when issued, will determine whether the exclusion of generative and agentic systems proves to be an interval or a settlement. Institutions that have built the capacity to evidence their own decisions in the meantime will be positioned either way; those that have read the contraction of the perimeter as a reduction in obligation will discover, in the examination room or in litigation, that the perimeter and the exposure were never the same boundary.

References

1. Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1), 99–120.
2. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671–732.
3. Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2022). Consumer-lending discrimination in the FinTech era. *Journal of Financial Economics*, 143(1), 30–56.
4. Berg, T., Burg, V., Gombović, A., & Puri, M. (2020). On the rise of FinTechs: Credit scoring using digital footprints. *The Review of Financial Studies*, 33(7), 2845–2897.
5. Board of Governors of the Federal Reserve System. (2011). SR 11-7: Guidance on model risk management. [Rescinded 2026]
6. Board of Governors of the Federal Reserve System. (2026). SR 26-2: Revised guidance on model risk management, April 17, 2026.
7. Board of Governors of the Federal Reserve System, Consumer Financial Protection Bureau, Federal Deposit Insurance Corporation, National Credit Union Administration, & Office of the Comptroller of the Currency. (2021). Request for information and comment on financial institutions' use of artificial intelligence, including machine learning. 86 Fed. Reg. 16837.
8. Bowman, M. W. (2026, May 1). Artificial intelligence in the financial system [Speech]. Board of Governors of the Federal Reserve System.
9. Bracke, P., Datta, A., Jung, C., & Sen, S. (2019). Machine learning explainability in finance: An application to default risk analysis (Staff Working Paper No. 816). Bank of England.
10. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
11. Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics*, 57(1), 203–216.
12. Colorado General Assembly. (2026). Senate Bill 26-189: Automated decision-making technology. [Effective January 1, 2027]
13. Consumer Financial Protection Bureau. (2022). Circular 2022-03: Adverse action notification requirements in connection with credit decisions based on complex algorithms.
14. Consumer Financial Protection Bureau. (2023). Circular 2023-03: Adverse action notification requirements and the proper use of the CFPB's sample forms provided in Regulation B.
15. Consumer Financial Protection Bureau. (2025). Supervisory highlights, Winter 2025.
16. Consumer Financial Protection Bureau. (2026). Circular 2026-03: Adverse action requirements for creditors using machine-learning underwriting models, May 5, 2026.
17. Consumer Financial Protection Bureau, Civil Rights Division of the U.S. Department of Justice, Federal Trade Commission, & Equal Employment Opportunity Commission. (2023). Joint statement on enforcement efforts against discrimination and bias in automated systems.
18. Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
19. DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147–160.
20. Equal Credit Opportunity Act, 15 U.S.C. § 1691 et seq.; Regulation B, 12 C.F.R. Part 1002.
21. European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*.
22. Executive Order 14281 of April 23, 2025. Restoring equality of opportunity and meritocracy.

23. Federal Deposit Insurance Corporation. (2026). FIL-16-2026: Interagency guidance on model risk management.
24. Financial Stability Board. (2017). Artificial intelligence and machine learning in financial services: Market developments and financial stability implications.
25. Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., & Walther, A. (2022). Predictably unequal? The effects of machine learning on credit markets. *The Journal of Finance*, 77(1), 5–47.
26. Gambacorta, L., Huang, Y., Qiu, H., & Wang, J. (2024). How do machine learning and non-traditional data affect credit scoring? *Journal of Financial Stability*, 73, 101284.
27. Gillis, T. B. (2022). The input fallacy. *Minnesota Law Review*, 106(3), 1175–1263.
28. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
29. Interagency guidance on third-party relationships: Risk management. (2023). 88 Fed. Reg. 37920.
30. Jagtiani, J., & Lemieux, C. (2019). The roles of alternative data and machine learning in fintech lending. *Financial Management*, 48(4), 1009–1029.
31. Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11), 2767–2787.
32. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
33. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229.
34. National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce.
35. Office of the Comptroller of the Currency. (2011). Bulletin 2011-12: Supervisory guidance on model risk management. [Rescinded 2026]
36. Office of the Comptroller of the Currency. (2025). Bulletin 2025-16: Fair lending examination procedures.
37. Office of the Comptroller of the Currency. (2026a). Bulletin 2026-13: Model risk management - Revised guidance, April 17, 2026.
38. Office of the Comptroller of the Currency. (2026b). Semiannual risk perspective, Spring 2026.
39. Power, M. (1997). *The audit society: Rituals of verification*. Oxford University Press.
40. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.
41. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215.
42. Scott, W. R. (2014). *Institutions and organizations: Ideas, interests, and identities* (4th ed.). SAGE Publications.
43. Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509–533