THE COMPUTERTECH
(An International Peer Review Journal)
VOLUME 12; ISSUE 2 (JULY-DEC); (2026)
WEBSITE: THE COMPUTERTECH

Change Control for Adaptive Machine-Learning Credit Models: A Tiered Approval Framework for Recalibration, Retraining and Redevelopment in Community Lending

Bwalya Marković [1], Tembo Petrović [2]
[1] University of Khartoum, SUDAN
[2] Birzeit University, PALESTINE

Abstract
Adaptive machine-learning credit models promise to keep early-warning systems accurate as economic conditions change, but every update is also a model change that supervisors expect to be controlled, validated and documented. Recent work on model risk management for adaptive credit early-warning systems in supervised small-business lending has argued that machine-learning recalibration must be aligned with federal model-risk guidance and the NIST AI Risk Management Framework [1]. Community banks and credit unions face this challenge with small model-risk teams, so a change-control process that treats every update as a full redevelopment is impractical, while one that treats every update as routine is unsafe. This article proposes a tiered change-control framework for adaptive credit models. It classifies changes into three tiers through a simple decision tree, introduces a pre-approved recalibration envelope within which routine updates may proceed with light-touch review, scales documentation, testing, validation and approval to each tier, and defines a six-stage lifecycle with approval gates. The framework is mapped to supervisory model-risk guidance and to the govern, map, measure and manage functions of the NIST AI RMF, and it draws on the predetermined change control plans now used for adaptive medical AI. An illustrative stress scenario shows how the tiers allocate review effort. We then test the envelope empirically on real SBA small-business loans, applying its trigger, bounds and automated tests to 144 monitoring scenarios. Gradual shifts in sector mix never breached the trigger (observed-to-expected defaults 0.97 to 1.20). Under credit-cycle shifts the trigger worked, but the materiality bounds escalated 36 recalibrations to Tier 2, of which 32 needed nothing more than recalibration; no scenario required retraining. The envelope bounds should therefore distinguish large but correct recalibrations from risky ones. The framework is intended to let smaller lenders benefit from adaptive models without weakening model governance.

Keywords: model risk management; change control; adaptive machine learning; credit early warning; recalibration; community banks; credit unions; NIST AI RMF

Introduction

Credit early-warning models for small-business loans lose accuracy when economic conditions change. Machine-learning methods make it technically easy to update such models frequently, and adaptive approaches can retrain or recalibrate them automatically as new outcomes arrive [2][3].

From a governance perspective, however, each update is a change to a model used in credit decisions. US supervisory guidance expects model changes to be subject to development standards, validation and approval commensurate with their risk [4][5].

The tension between adaptivity and control is the focus of recent work on model risk management for adaptive credit early-warning systems in supervised small-business lending [1], which argues for aligning machine-learning recalibration with federal model-risk guidance and the NIST AI Risk Management Framework [6]. This article develops one practical element of that alignment, a change-control process, and tests its recalibration envelope on public small-business loan data. The question is especially pressing for community banks and credit unions, whose model-risk functions are small. If every update requires a full validation, adaptive models become impractical; if updates bypass validation, the institution takes on unmanaged risk.

We propose a tiered framework that matches the depth of review to the materiality of each change. Its components are a taxonomy of changes, a decision tree assigning each change to one of three tiers, a pre-approved recalibration envelope, a six-stage lifecycle with approval gates, tier-specific documentation and testing requirements, governance roles, and a mapping to supervisory and NIST expectations.

Background and Related Work

Supervisory model risk guidance defines a model broadly and expects sound development, independent validation and governance throughout its life, including when it is changed [4]. The OCC's handbook elaborates these expectations for examiners, describing change management, ongoing monitoring and the scaling of validation to model risk [5]. Validation guidance for internal rating systems similarly distinguishes between monitoring that a model is performing as intended and re-validation following material change [7]. Interagency regulators have also sought information on how institutions manage the specific risks of artificial intelligence, including dynamic updating and explainability [8].

Machine-learning engineering research highlights why change control is harder for learned systems. Production machine-learning systems accumulate hidden technical debt through data dependencies, feedback loops and configuration complexity [9], and case studies of deployment show that data management, testing and monitoring account for much of the practical difficulty [10][11]. Rubrics for production readiness recommend explicit tests for data, models, infrastructure and monitoring [12]. Documentation practices such as model cards record a model's intended use, performance and limitations [13], and internal algorithmic audit frameworks describe how organizations can trace decisions about AI systems from design to deployment [14].

Outside banking, regulators of medical devices have confronted the same problem. The US Food and Drug Administration's guidance on predetermined change control plans allows manufacturers of AI-enabled device software to describe in advance the modifications they intend to make, the methods for making them and the assessment of their impact, so that changes within the plan need not each be submitted anew [15]. This idea of pre-specifying acceptable changes informs the recalibration envelope proposed below.

Why Adaptive Models Strain Traditional Change Control

Traditional change control assumes that models change rarely and deliberately. Adaptive credit models invert both assumptions. They may change quarterly or monthly, and some changes, such as recalibration of the score-to-probability mapping, are triggered automatically by monitoring rules rather than initiated by a developer. Changes also arise from sources other than the model itself: a new data feed, a revised threshold, a software library upgrade or a decision to use the model for a new product. Table 1 classifies the main sources of change.

Table 1. Sources of change in adaptive credit early-warning models.

Source	Examples	Typical risk
Model structure	New features; new algorithm; changed target definition	Behavior changes in ways hard to predict
Fitted parameters	Retraining on recent data; recalibration of probabilities	Shifts in risk estimates across the portfolio
Decision thresholds	New watch-list cut-off; changed alert rules	More or fewer borrowers flagged
Data and infrastructure	New vendor feed; changed field mapping; library upgrade	Silent errors in inputs or scores [9]
Use	Extending the model to a new product or segment	Model applied outside its validated scope

Tiered Change Classification

Figure 1 presents a decision tree that assigns each proposed change to a tier. Changes to the model's inputs or structure are Tier 1, major. Changes to fitted parameters, including retraining with the same structure and recalibration, are Tier 2, moderate. Changes to thresholds, code or data feeds that leave the model's logic unchanged are Tier 3, minor. Anything that alters none of these is recorded but is not treated as a model change.

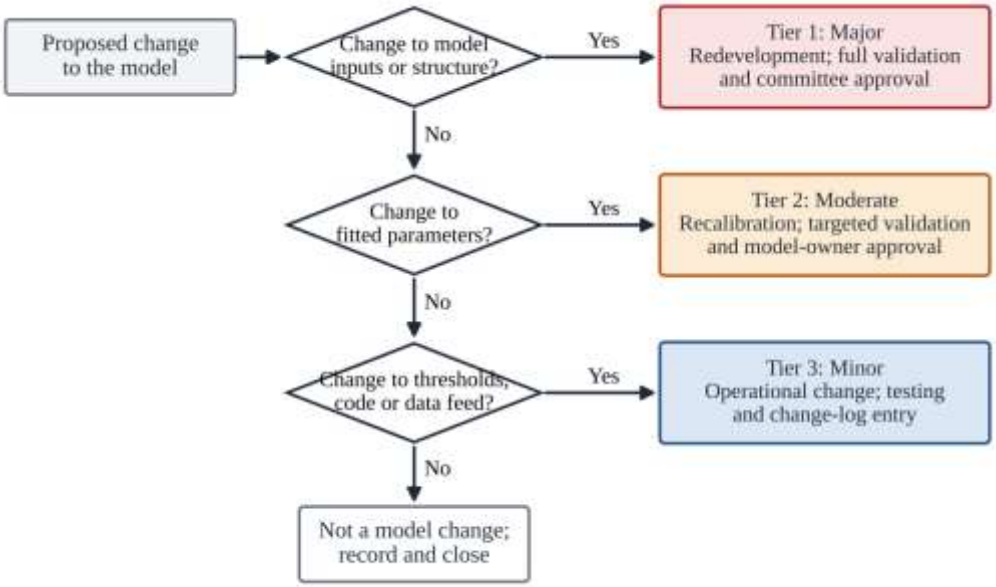


Figure 1. Decision tree for classifying proposed changes to an adaptive credit model [authors' design].

The tree gives a default tier. Table 2 lists materiality factors that can raise a change by one tier, so that, for example, a recalibration that would move a large share of the portfolio across an action threshold is treated as major rather than moderate. These escalation factors reflect the supervisory principle that the rigour of control should match the potential impact of the model [4][5].

Table 2. Materiality factors that raise a change by one tier.

Factor	**Illustrative test**	**Effect**
Portfolio impact	More than 10% of borrowers change watch-list status	Raise one tier
Risk-estimate shift	Portfolio average predicted default rate changes by more than 25% relative	Raise one tier
Fair-lending sensitivity	Change affects groups differently at equal risk	Raise one tier; add fairness review
New use or population	Model extended beyond its validated scope	Always Tier 1
Repeated changes	Third Tier 2 change within twelve months	Raise one tier

The Pre-Approved Recalibration Envelope

Recalibration is the most frequent change an adaptive credit model undergoes, and it is also the most predictable. Recalibration adjusts the mapping from scores to probabilities without changing the ranking of borrowers [16]. Following the logic of predetermined change control plans [15], the

framework allows an institution to approve in advance an envelope within which recalibration may proceed under a simplified Tier 3 process. A recalibration falling outside the envelope reverts to Tier 2. Table 3 sets out the elements of an envelope.

Table 3. Elements of a pre-approved recalibration envelope.

Element	**Content approved in advance**
Trigger	Monitoring rule that initiates recalibration, such as observed-to-expected defaults outside 0.8–1.2
Method	Specified recalibration method and data window, for example logistic recalibration on the latest two labelled quarters
Bounds	Maximum permitted change in intercept and slope, and in the portfolio average predicted default rate
Tests	Automated checks: rank order unchanged; calibration improved on hold-out data; fairness metrics within limits
Reporting	Change-log entry and notification to the model risk function within five working days
Review	Annual review of all envelope recalibrations as part of model validation

The envelope makes routine adaptation fast while keeping it visible and bounded. It also creates a clear signal: repeated recalibrations, or a recalibration that breaches the bounds, indicate that the underlying relationships may have changed in ways that only retraining or redevelopment can address, consistent with the escalation logic advocated in [1].

Lifecycle and Approval Gates

Every change, whatever its tier, passes through the same six stages shown in Figure 2: request and classification, development and testing, independent validation, approval, implementation with a parallel run, and post-change monitoring. Approval gates separate the stages. What changes by tier is the depth of each stage, not whether it occurs, so that the change log remains complete and auditable [14].

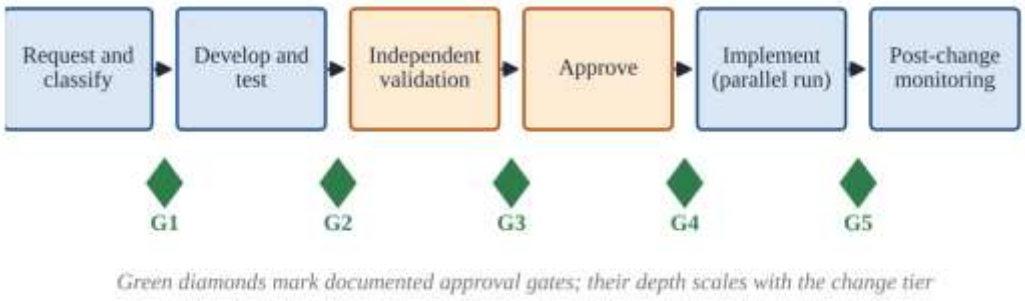


Figure 2. Six-stage change lifecycle with approval gates [authors' design].

Implementation with a parallel run is especially valuable for adaptive credit models. Running the old and new versions side by side for one reporting cycle shows how many borrowers would change status and whether the change behaves as expected, before the new version drives decisions. Testing practices from machine-learning engineering, such as checks on input distributions, prediction stability and serving consistency, can be automated for this stage [12].

Documentation and Rigour by Tier

Figure 3 summarizes how the depth of documentation, testing, validation, approval and parallel running scales with tier. Tier 1 changes receive full treatment equivalent to a new model. Tier 2 changes receive targeted validation focused on the parameters that changed. Tier 3 changes receive light documentation, automated tests and peer review rather than independent validation.

	Documentation	Testing	Independent validation	Approval level	Parallel run
Tier 1: Major	Full	Full	Full	Full	Full
Tier 2: Moderate	Targeted	Targeted	Targeted	Targeted	Light
Tier 3: Minor	Light	Light	Peer review	Light	Not required

Figure 3. Depth of control activities by change tier [authors' design].

Each change is recorded in a change log whose fields extend the documentation practices of model cards [13]. Table 4 lists the minimum fields. A complete log allows validators and examiners to reconstruct the history of the model and to see the cumulative effect of many small changes, which may be material even when no single change is.

Table 4. Minimum fields for each change-log entry.

Field	Content
Identifier and date	Unique change identifier; model version before and after
Classification	Tier from the decision tree; materiality factors applied
Rationale	Monitoring evidence or business reason that prompted the change
Description	What changed: data, parameters, thresholds, code or use
Testing results	Pre- and post-change performance, stability and fairness metrics
Validation and approval	Reviewer, approver, date and any conditions
Parallel-run outcome	Number of borrowers changing status; issues found
Rollback plan	How to restore the previous version if problems emerge

Governance Roles

Clear roles are essential when a small institution handles many changes. Table 5 assigns responsibilities across model owners, developers, the model risk function, a model risk committee and internal audit. Where the model is supplied by a vendor, contracts should give the institution the information needed to classify and validate the vendor's changes, in line with third-party risk guidance [17].

Table 5. Roles in the tiered change-control framework.

Role	Responsibilities	Approves
Model owner (business line)	Requests changes; confirms business need; monitors use	Tier 3 changes
Model developer or vendor	Builds and tests changes; prepares documentation	None
Model risk function	Classifies changes; validates Tier 1 and 2; maintains envelope	Tier 2 changes
Model risk committee	Sets policy, envelope bounds and materiality factors	Tier 1 changes
Internal audit	Tests that the process operates as designed	None (assurance)

Alignment with Supervisory Guidance and the NIST AI RMF

Table 6 maps the framework to supervisory model-risk expectations and to the four functions of the NIST AI Risk Management Framework. The mapping shows that the framework does not add new obligations so much as organize existing ones for the specific case of adaptive models, in the spirit of the alignment proposed in [1].

Table 6. Mapping of the framework to supervisory guidance and the NIST AI RMF.

Framework element	**Supervisory expectation [4][5]**	**NIST AI RMF function [6]**
Change policy, tiers and envelope	Policies and procedures for model changes	Govern
Classification and materiality factors	Rigour commensurate with model risk	Map
Testing, parallel run and monitoring	Outcomes analysis and ongoing monitoring	Measure
Approval gates and rollback plans	Controls over model use and change	Manage
Change log and audit	Documentation and effective challenge	Govern

Consider a community bank whose small-business early-warning model begins to under-predict defaults as a regional downturn develops. In the first quarter of stress, monitoring shows observed defaults 25 percent above expectation. Because the recalibration envelope permits adjustments when the ratio leaves the 0.8–1.2 band, the model owner applies a pre-specified recalibration as a Tier 3 change; automated tests confirm that rank order is unchanged and calibration improves, and the change is logged and reported.

Two quarters later, the score distribution has shifted markedly and a second recalibration would exceed the envelope's bounds. The change is reclassified as Tier 2 and the model risk function performs a targeted validation. When analysis shows that the effect of credit-line utilization on default has strengthened, the developer proposes adding a new utilization-trend feature. That proposal changes the model's inputs and is therefore Tier 1: it receives full validation, committee approval and a parallel run before deployment. In this way the framework allocates scarce review effort where the risk is greatest, while keeping routine adaptation quick.

Empirical Test of the Recalibration Envelope

The envelope in Table 3 makes testable claims: its trigger should fire when calibration has drifted, recalibration inside its bounds should be adequate, and escalation should be reserved for changes that recalibration cannot handle. We tested these claims on 80,000 public SBA-guaranteed small-business loans[18] used for development, with further real loans from the same file used to build monitoring populations. The champion was a LightGBM distress model[19] trained on four sector portfolios (retail, professional services, manufacturing and health care) using origination-time features (development AUC 0.758). The envelope was implemented exactly as in Table 3: trigger when observed-to-expected (O/E) defaults on the latest labelled window of 4,000 loans fall outside 0.8 to 1.2; logistic recalibration[16]; bounds of 0.5 on the intercept, 0.8 to 1.25 on the slope and 25 percent on the change in average predicted default; automated tests that rank order is preserved and calibration improves on a hold-out of 8,000 loans; and escalation by one tier if more than 10 percent of borrowers change watch-list status (Table 2). Each decision was compared with what was actually needed, judged against a model retrained with the new labels.

Scenario Set 1: Gradual Change in Portfolio Mix

In the first set, monitoring populations mixed loans from four new sectors (construction, other services, wholesale trade, and accommodation and food) into the original book at shares from 0 to 100 percent (84 scenarios). O/E stayed between 0.97 and 1.20, so the trigger never fired, and the retrained benchmark confirmed that no action was needed in 82 of the 84 cases. Discrimination fell only slightly as the new-sector share rose (AUC from 0.744 to 0.731). The envelope correctly stayed quiet, although even a full move into new sectors would count as a "new population" under Table 2 and should be classed as Tier 1 regardless of what the calibration trigger shows.

Scenario Set 2: Credit-Cycle Shifts

In the second set, monitoring populations were drawn from real loans resampled to default rates from 0.6 to 2.0 times the development rate of 19.1 percent, alone or combined with a 50 percent new-sector mix (60 scenarios). Table 7 summarises the envelope's decisions.

Table 7. Recalibration-envelope decisions under credit-cycle scenarios (real SBA loans; 10 scenarios per default-rate level).

Default rate vs development	Mean O/E	Trigger fired	Recalibrated in envelope	Escalated	Recalibration was sufficient	Watch-list churn	Brier: champion	Brier: recalibrated	Brier: retrained
0.60x	0.64	100%	0%	100%	100%	6.0%	0.098	0.092	0.096
0.80x	0.81	40%	40%	0%	30%	3.7%	0.117	0.115	0.116
1.00x	0.97	0%	0%	0%	0%	1.4%	0.137	0.137	0.137
1.25x	1.22	60%	0%	60%	40%	15.5%	0.158	0.157	0.158
1.50x	1.39	100%	0%	100%	100%	28.1%	0.182	0.174	0.179
2.00x	1.75	100%	0%	100%	100%	49.9%	0.228	0.196	0.216

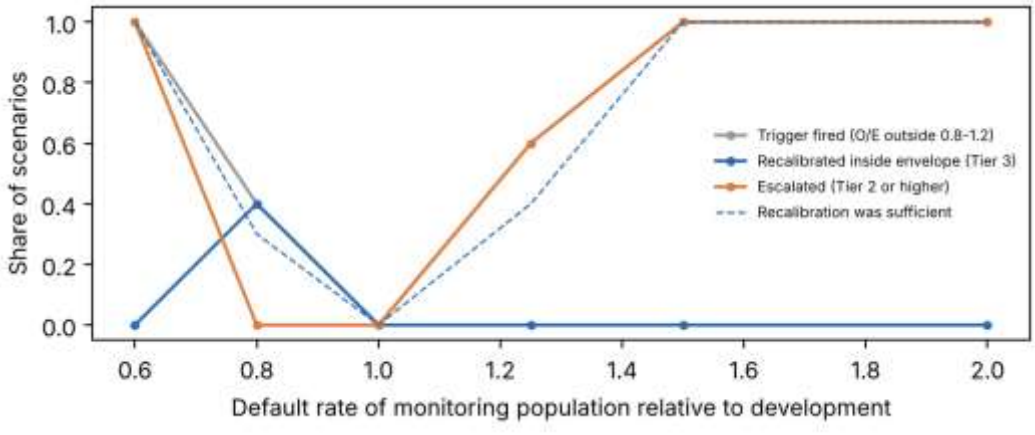


Figure 4. Share of credit-cycle scenarios triggering the envelope, recalibrated inside it, or escalated, against the share in which recalibration was sufficient [authors' analysis].

The trigger behaved as intended, firing in every scenario with a default-rate shift of 1.5 times or 0.6 times and in none at the development rate. The bounds did not. Of the 36 scenarios the envelope escalated to Tier 2, recalibration alone would have been sufficient in 32: the recalibrated model matched or beat the retrained model on hold-out Brier score without any loss in ranking. In the 2.0x scenarios, for example, recalibration lowered the Brier score from 0.228 to 0.196, better than retraining (0.216). Escalations were driven by the change in average predicted default (32 scenarios), watch-list churn (26) and the intercept bound (22); the slope bound was breached only 1 time. Only 4 scenarios were handled inside the envelope, 2 scenarios where recalibration would have helped went untriggered, and none of the 60 scenarios needed retraining.

The result is a specific lesson for the design in Table 3. Under a credit-cycle shift, a correct recalibration is necessarily large: it moves the average predicted default and many borrowers' watch-list status, because that is what the portfolio's risk has done. Bounds on those quantities therefore escalate exactly the recalibrations that are most needed and least risky, sending routine work to Tier 2 review. A better envelope would bound the slope, which signals a change in ranking, and require calibration improvement on hold-out data, while treating large intercept shifts and watch-list churn as reporting items for the model-risk function rather than as escalation triggers. Portfolio impact still matters for fair-lending and capital purposes, but it is evidence of what the change does, not of whether the change is valid.

Discussion

The framework reconciles two goods that are often presented as competing: the responsiveness of adaptive machine learning and the discipline of model risk management. Its central device, matching review depth to the materiality of each change, is already implicit in supervisory guidance [4][5]. What the framework adds is a concrete and repeatable way to apply it to frequent, partly automated updates. The pre-approved envelope, borrowed from medical-device regulation [15], is particularly useful for small institutions, because it converts a recurring decision into a one-time policy choice that can be reviewed annually.

Two cautions apply. First, the envelope must be bounded tightly enough that repeated small changes cannot accumulate into a large unreviewed one; the repeated-change materiality factor and the annual review address this. Second, fairness must be tested at every tier, since even a simple recalibration can affect groups differently when their score distributions differ.

Conclusion

Adaptive machine-learning credit models need change control that is as adaptive as they are. This article has proposed a tiered framework that classifies changes through a simple decision tree, permits routine recalibration within a pre-approved envelope, scales documentation, testing, validation and approval to each tier, and maps every element to supervisory guidance and the NIST AI RMF. Tested on 144 scenarios built from real SBA loans, the envelope's trigger worked, but its bounds on portfolio impact escalated 32 recalibrations that were sufficient on their own, which argues for bounding the slope rather than the size of the correction. The framework allows community lenders to keep early-warning models accurate as conditions change while maintaining the governance that supervisors, boards and borrowers expect.

References

[1] Karim, N. B., Kovac, T., & Song-Chun, Z. (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. *International Journal of Acta Informatica*, 4(1), 80-94.

[2] Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys, 46*(4), Article 44.

[3] Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., & Zhang, G. (2019). Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering, 31*(12), 2346–2363.

[4] Board of Governors of the Federal Reserve System, & Office of the Comptroller of the Currency. (2011). *Supervisory guidance on model risk management* (SR Letter 11-7; OCC Bulletin 2011-12).

[5] Office of the Comptroller of the Currency. (2021). *Comptroller's handbook: Model risk management* (Version 1.0).

[6] 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.

[7] Basel Committee on Banking Supervision. (2005). *Studies on the validation of internal rating systems* (Working Paper No. 14). Bank for International Settlements.

[8] Board of Governors of the Federal Reserve System, Bureau of Consumer Financial Protection, 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. *Federal Register, 86*, 16837.

[9] Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. In *Advances in Neural Information Processing Systems* (Vol. 28, pp. 2503–2511).

[10] Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software engineering for machine learning: A case study. In *Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice* (pp. 291–300). IEEE.

[11] Paleyes, A., Urma, R.-G., & Lawrence, N. D. (2022). Challenges in deploying machine learning: A survey of case studies. *ACM Computing Surveys, 55*(6), Article 114.

[12] Breck, E., Cai, S., Nielsen, E., Salib, M., & Sculley, D. (2017). The ML test score: A rubric for ML production readiness and technical debt reduction. In *Proceedings of the 2017 IEEE International Conference on Big Data* (pp. 1123–1132). IEEE.

[13] 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. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220–229). ACM.

[14] Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33–44). ACM.

[15] U.S. Food and Drug Administration. (2024). *Marketing submission recommendations for a predetermined change control plan for artificial intelligence-enabled device software functions: Guidance for industry and Food and Drug Administration staff*.

[16] Platt, J. C. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In A. J. Smola, P. Bartlett, B. Schölkopf, & D. Schuurmans (Eds.), *Advances in large margin classifiers* (pp. 61–74). MIT Press.

[17] Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation, & Office of the Comptroller of the Currency. (2023). Interagency guidance on third-party relationships: Risk management. *Federal Register*, *88*, 37920.

[18] Li, M., Mickel, A., & Taylor, S. (2018). "Should this loan be approved or denied?": A large dataset with class assignment guidelines. *Journal of Statistics Education*, 26(1), 55-66.

[19] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems 30* (pp. 3146-3154).