PASCHAL ALUMONA

Proposal and Comment Information

Title: Anti-Money Laundering and Countering the Financing of Terrorism Programs, R-1835

Comment ID: FR-0000-0133-02-C17

Submitter Information

Name: Paschal Alumona

Submitted Date: 08/11/2026

Please see the attached PDF for my full comment on Docket No. R-1835; RIN 7100-AG78 (Anti-Money Laundering and Countering the Financing of Terrorism Programs, Notice of Proposed Rulemaking, 91 FR 42363, July 9, 2026).

I am an independent researcher and a co-author of published open-access studies on the evaluation of interpretable machine-learning models for fraud detection in financial transaction systems. I submit this comment in my personal capacity; the views expressed are my own.

The attached comment addresses one subject: how machine-learning transaction monitoring can be evaluated and documented so that both institutions and examiners can rely on it. In brief, I respectfully recommend that the Board (1) treat documented pre-deployment error-cost and threshold evaluation as favorable evidence that an innovative monitoring approach is reasonably designed; (2) treat retained explainability (feature-attribution) documentation as what makes such an approach examinable; and (3) preserve room for documented, interpretable approaches at small Board-supervised institutions, avoiding any implication that innovation means model complexity or single-metric performance claims.

Respectfully submitted,
Paschal Tochukwu Alumona
Independent researcher

August 11, 2026

Board of Governors of the Federal Reserve System

Attention: Comments, Docket No. R-1835; RIN 7100-AG78

Re: Anti-Money Laundering and Countering the Financing of Terrorism Programs: Notice of Proposed Rulemaking, 91 FR 42363 (July 9, 2026); Docket No. R-1835; RIN 7100-AG78

Ladies and Gentlemen:

I am an independent researcher and a co-author of published open-access studies on the evaluation of interpretable machine-learning models for fraud detection in financial transaction systems. I submit this comment in my personal capacity; the views expressed are my own. I write on one subject only: how machine-learning transaction monitoring can be evaluated and documented so that both institutions and examiners can rely on it. I offer no view on other aspects of the proposal.

1. The provisions this comment addresses

The preamble's discussion of innovative approaches to program requirements states: "Innovative approaches could involve machine learning, generative artificial intelligence (GenAI), digital identity, blockchain monitoring and analytics, or application programming interfaces (APIs)." The preamble also quotes the Anti-Money Laundering Act's stated purpose to "encourage technological innovation and the adoption of new technology by financial institutions to more effectively counter money laundering." This comment addresses how those provisions can be implemented so that machine-learning monitoring is adopted in a form institutions can defend and examiners can examine, particularly at small Board-supervised institutions.

2. Error-cost and threshold evaluation before deployment should be an examinable practice

A machine-learning transaction-monitoring model is not a single object with a single quality score. It is a family of operating points: at each alert threshold, the model trades one error class against the other. In monitoring terms, the two error classes carry different costs. A missed suspicious transaction is a compliance failure; a false alert consumes analyst time, and at volume it buries the true alerts. A model evaluation that reports one aggregate metric hides this trade; an evaluation that states it lets an institution choose, and defend, an operating point.

A published study I co-authored illustrates the point concretely. My co-authors and I trained and compared four supervised classifiers on the public PaySim synthetic mobile-money dataset, using stratified splitting and class weighting to address the dataset's extreme class imbalance (fraud is a rare event, so most metrics are misleading unless the imbalance is handled explicitly). On the held-out test set, as the paper reports, the Decision Tree model achieved Precision = 0.6835, Recall = 0.9696, F1 = 0.8018, and ROC-AUC = 0.9845. The Random Forest model achieved higher recall (0.9838) and a higher ROC-AUC (0.9990), but a precision of 0.1576, meaning most of the transactions it flagged were not fraudulent. Ranked on ROC-AUC alone, the Random Forest is the

"better" model. Priced in error costs, it would flood an alert queue with false positives that each require human review, a cost a large institution can absorb and a small one cannot, while the Decision Tree offers a precision-recall balance a compliance function can actually operate. These are the study's own reported test-set results on one public synthetic dataset; I present them as a worked illustration of an evaluation practice, not as a claim about any production system.

The practice this illustrates is documentable and examinable: before deployment, state the candidate models, the class-imbalance handling, the per-threshold false-positive/false-negative trade, and the reasoning for the operating point chosen. That documentation is something an examiner can read and test. I respectfully suggest that the final rule and any implementing guidance treat documented pre-deployment error-cost and threshold evaluation as favorable evidence that a machine-learning monitoring approach is part of a reasonably designed, risk-based program.

3. Explainability documentation is what makes machine-learning monitoring auditable

The second evaluation practice is feature-attribution documentation: identifying, and retaining a record of, which transaction features drive a model's alerts. In a separate published study I co-authored, a gradient-boosted (XGBoost) model was trained on a public card-transaction dataset of 284,807 transactions containing 492 confirmed frauds, roughly 0.17 percent of the data, again an extreme class imbalance that makes single-number metrics uninformative on their own. Evaluated with imbalance-aware metrics, the paper reports a ROC-AUC of 0.975, and its global feature-importance analysis identified the anonymized features designated V14, V10, and V4 as the dominant predictors. The specific features matter less than the practice: when the dominant decision factors of a monitoring model are computed and recorded, a compliance officer can review what the model relies on, justify or escalate a specific alert decision, and detect drift when the dominant factors change. Without that record, the model is a vendor's black box, and the institution is certifying a process it cannot explain.

I respectfully suggest that guidance implementing the innovation provisions recognize retained feature-attribution (model-explanation) documentation as the characteristic that makes machine-learning monitoring auditable: by the institution's own compliance staff first, and by examiners second.

4. Small institutions need evaluation protocols they can operate without model-risk teams

Both practices above run on modest resources: the two studies cited here used public datasets, open-source tooling, and computing at consumer scale. That matters for the proposal's coverage, which includes small Board-supervised institutions. Such institutions face the same monitoring obligations as the largest banks but typically employ no model-risk team. For them, the realistic path to the innovation the statute encourages is not the most complex model but a documented, interpretable one: a model whose error trade-offs were stated before deployment and whose decision factors a small compliance staff can review. If the final rule's innovation provisions are read in practice

to reward complexity or headline metrics, small institutions will face a choice between opaque vendor systems they cannot explain and no adoption at all.

Recommendation. In finalizing the rule and in any related guidance or examination procedures, I respectfully recommend that the Board: (1) treat documented pre-deployment error-cost and threshold evaluation as favorable evidence that an innovative monitoring approach is reasonably designed; (2) treat retained explainability (feature-attribution) documentation as what makes such an approach examinable; and (3) preserve room for documented, interpretable approaches at small institutions, avoiding any implication that innovation means model complexity or single-metric performance claims.

5. Published studies cited

1. "Fraud Detection in Financial Transactions Using Machine Learning: Insights from the PaySim Mobile Money Dataset," *World Journal of Advanced Research and Reviews*, 2025, 28(03), 382-392. DOI: 10.30574/wjarr.2025.28.3.4058. I am the first and corresponding author of this study.

2. "An Explainable XGBoost Framework for Detecting Fraudulent Financial Transactions," *Journal of Scientific Research and Reports*, 2025, 31(12), 244-255. DOI: 10.9734/jsrr/2025/v31i123769. I am a co-author of this study.

Both are open-access publications; the figures quoted above are the papers' own reported results, subject to the class-imbalance caveats stated in each paper and to the limits of single-dataset evaluation.

Thank you for considering this comment.

Respectfully submitted,
Paschal Tochukwu Alumona
Independent researcher
Chicago, Illinois