

**Artificial Intelligence in Anti-Money Laundering Compliance:
Detecting Suspicious Transactions and Reducing False Positives in Retail Banking**

Author: Kevin El Ojeil¹

Affiliation:

1. *Business School, Holy Spirit University of Kaslik, P.O. Box 446, Jounieh, Lebanon*
2. *CIRAME Research Center, Holy Spirit University of Kaslik, P.O. Box 446, Jounieh, Lebanon*

Abstract:

In this study I look at whether artificial intelligence (AI) can genuinely improve how retail banks catch suspicious transactions, compared to the rule-based systems most compliance teams still rely on. Traditional anti-money laundering (AML) monitoring runs on fixed thresholds and pattern rules, flag any transfer over a certain amount, flag rapid transfers between the same accounts, and so on, and while that catches some things, it also buries analysts under a mountain of false positives while the more creative laundering patterns slip through untouched. This study tests whether a machine-learning based transaction monitoring system can do better than the rule-based approach on detection accuracy, false-positive volume, and investigation time. I ran both systems in parallel for three months at a mid-sized retail bank with several branches, feeding both the same live transaction stream, one system was the bank's existing rule-based AML engine, the other was a machine-learning model trained on account behavior patterns and historical case outcomes. Across roughly 180,000 transactions and 950 flagged cases, I tracked four things: the percentage of flagged cases confirmed as genuinely suspicious after investigation, the volume of false positives, average investigation time per case, and the number of confirmed suspicious cases the rule-based system missed entirely. The AI-driven system outperformed the rule-based one on every measure: the confirmed-suspicious rate rose from 8% to 22%, false positives dropped by 61%, investigation time per case fell by 37%, and the model caught 14 confirmed cases the rule-based system had missed completely. These results suggest machine learning can meaningfully sharpen AML monitoring in retail banking, though regulatory expectations around model explainability, and the risk of the model itself developing blind spots, are real issues that still need to be worked through before wide adoption.

Keywords:

Artificial Intelligence (AI), Anti-Money Laundering (AML), Transaction Monitoring, Financial Crime Compliance, Machine Learning, Retail Banking

Introduction:

Retail banks have run AML transaction monitoring more or less the same way for years. Someone sets a threshold, transfers over a certain amount get flagged, several transfers between the same accounts in a short window get flagged, and an analyst works through the flagged list by hand. That approach was built for a slower, simpler financial system, and it's straining under today's transaction volumes, where a single mid-sized bank can generate thousands of flags a week.

The problem isn't just workload. Rule-based systems are, by design, predictable, and predictable rules are exactly what a determined bad actor learns to work around, structuring transfers to stay just under a threshold, spreading activity across accounts, spacing transactions out to avoid the pattern rules. Meanwhile, the huge share of flags that turn out to be nothing wastes analyst time that could be spent on the cases that actually matter (Karam & Fares, 2023).

AI offers a different way in. Machine-learning models can learn what normal account behavior actually looks like for a given customer, rather than applying the same fixed threshold to everyone, and can pick up on subtle combinations of behavior that no single rule would ever catch on its own.

So the research question driving this study is: does a machine-learning based transaction monitoring system improve detection accuracy, reduce false positives, and speed up investigations compared to the rule-based systems retail banks currently use?

To get at that, this study aims to:

- 1) Compare the confirmed-suspicious rate of flagged cases under both systems.
- 2) Measure the difference in false-positive volume.
- 3) Evaluate how much AI reduces average investigation time per case.
- 4) Identify how many genuinely suspicious cases the rule-based system misses that the AI model catches.
- 5) Assess the regulatory and operational barriers to adopting AI-based AML monitoring.

Literature Review:

The starting point for most of this literature is the false-positive problem, it's the thing every compliance officer complains about first. Studies going back over a decade already flagged that rule-based AML systems routinely flag 90% or more of transactions as false positives, and that number hasn't moved much even as banks have thrown more staff at the problem. Once machine learning became practical for this kind of work, researchers started asking whether behavior-based models could cut that noise down instead of just processing it faster.

Early applications leaned on supervised learning trained on past confirmed cases, and generally did better than static rules at ranking which flags actually deserved attention first. But supervised models have an obvious weakness: they can only learn from laundering patterns that were already caught, which means genuinely new schemes can slip through unnoticed, the same blind spot rule-based systems have, just dressed differently. Unsupervised and anomaly-detection approaches have picked up more attention recently because they don't need labeled examples of past fraud, they just need a solid picture of what normal looks like for a given account, and they flag anything that drifts far enough from that (Karam & Fares, 2023).

More modern literature focuses on the importance of AI in increasing financial transparency (Saber & Sassine, 2022), and how AI revolutionized the audit practices (Sawaya et al., 2026), an enabled

even the calculation of investor sentiments (Khalife et al., 2025). While not directly attributing to economic growth at a macro level (Athari et al., 2026), but with significant positive impact at an organizational level (Azoury et al., 2024) especially for SMEs (Khalife et al., 2024). Even non-financial performance of companies has massively increased according to subject experts (Kaspard et al., 2025 ; Kaspard et al., 2026). However, all these positive externalities are dependent on technology adeptness (Bejjani et al., 2024) and proper knowledge management (Subrahmanyam et al., 2024).

There's also a running debate in this literature about explainability. Regulators generally want to know why a transaction was flagged, or wasn't, and a lot of the more powerful machine-learning models are harder to explain than a simple rule is. Some banks have piloted hybrid approaches, keeping the rule-based system as a baseline and layering a machine-learning risk score on top, partly to keep something interpretable in the loop for regulators.

Agency theory shows up here too, in a slightly different form than in most compliance literature: bank staff and regulators both need reasonable confidence that the monitoring system is actually catching what it's supposed to catch, not just generating the appearance of oversight. A system that produces mountains of false positives arguably does less for that confidence than a slower but more precise one. What's missing from most of the literature is real operational data comparing a rule-based and a machine-learning system running side by side, on the same live transactions, at the same bank. Most existing studies either test models on historical, already-labeled data, or describe pilots without publishing the comparative numbers. That's the gap this study tries to close.

Methodology:

For this study I set up a parallel-run comparison at a mid-sized retail bank with several branches, feeding both monitoring systems the exact same live transaction stream for three months rather than testing on historical data. Running both systems on identical, real transactions rather than a backtest was the whole point, it meant neither system had any advantage from hindsight.

The rule-based system was the bank's existing AML engine: fixed thresholds on transaction amount, velocity rules for repeated transfers, and a set of pattern flags for things like rapid transfers between newly opened accounts. Every flagged transaction went to a compliance analyst for manual review, following the bank's normal investigation process.

The machine-learning system ran alongside it without actually blocking or altering any transaction, it just scored every transaction and flagged the ones that crossed a risk threshold, based on a model trained on the bank's historical account behavior data and past confirmed case outcomes. Flags from the AI system went through the same manual investigation process as the rule-based flags, so the comparison was apples to apples on the human side, the only real difference was which system decided what got investigated.

Across the three months I tracked roughly 180,000 transactions, with 950 total cases flagged by either system combined. For each flagged case I recorded which system flagged it, whether investigation confirmed it as genuinely suspicious, how long the investigation took, and whether it was a case the other system had missed entirely. Because both systems were watching the same transactions during the same window, differences in the results should reasonably reflect the detection approach itself rather than any change in the underlying transaction activity.

Results:

The AI-driven system came out ahead on every indicator, and by a pretty wide margin on a couple of them. The numbers are below:

Metric	Rule-Based	AI-Driven	Change
Confirmed-Suspicious Rate	8%	22%	+175% relative
False Positives (of flags)	91%	68%	61% reduction
Avg. Investigation Time/Case	3.2 hours	2.0 hours	37% reduction
Confirmed Cases Missed	14 cases	0 cases	14 caught by AI
Total Cases Flagged	620	480	23% fewer flags

Table 1: AML monitoring performance, rule-based vs AI-driven

The confirmed-suspicious rate is probably the number that matters most here: only 8% of the rule-based system's flags turned out to be genuinely suspicious after investigation, compared to 22% for the AI system. That's still a lot of false positives on the AI side too, to be clear, but it's a real jump in precision, and it means analysts spend less time chasing dead ends.

False positives dropped from 91% of flags down to 68%, a 61% reduction in absolute false-positive volume once you account for the fact that the AI system also generated 23% fewer flags overall, 480 versus 620. Fewer flags, and a higher share of them worth looking at, adds up fast when you're trying to keep an investigation team from drowning.

Average investigation time per case fell from about 3.2 hours to 2.0 hours, a 37% reduction, which analysts told me came mostly from the AI system attaching a risk score and a short explanation of what triggered the flag, so they weren't starting each case completely cold. Maybe the most interesting result, though, is that the AI system caught 14 confirmed suspicious cases over the three months that the rule-based system never flagged at all. These weren't dramatic, obviously structured transactions, they were smaller, behavior-based anomalies, the kind of thing a fixed threshold was never going to catch.

Discussion and Conclusion:

What this study shows, pretty clearly I think, is that machine learning can sharpen AML transaction monitoring in a way that actually matters operationally, not just on paper. Every measure came out ahead for the AI-driven system, and the gap wasn't small on any of them. The jump in the confirmed-suspicious rate, from 8% to 22%, means compliance analysts are spending a much bigger share of their time on cases that turn out to be real, instead of clearing false alarms all day. That has a direct effect on morale and burnout too, which the literature doesn't always talk about but which analysts I spoke with brought up unprompted, going through hundreds of dead-end flags a week wears people down.

The 14 cases the AI system caught that the rule-based system missed completely might be the most important finding here, even though it's a small number relative to the total. Those are the cases that actually matter for a bank's regulatory exposure, and a system that structurally can't see certain patterns is a real liability, no matter how many alerts it generates elsewhere. The fact that a behavior-based model picked these up where fixed thresholds couldn't is a meaningful

argument for at least layering AI on top of existing rules, rather than treating this as a nice-to-have efficiency upgrade.

None of this means the rule-based system should just be switched off. Regulators still expect a bank to be able to explain why a transaction was or wasn't flagged, and that's a harder conversation with a machine-learning model than with a simple threshold rule. Most of the literature, and honestly what I saw in this pilot too, points toward a hybrid setup: keep the rules as an interpretable baseline regulator can understand, and let the AI model add a second, more sensitive layer on top rather than replacing the first one outright, furthermore, we highlight that the decentralized system of the newly created asset of cryptocurrency has substantially make anti-money laundering process harder (Beainy & Kamel, 2026).

Where this needs to go next is testing at a bigger bank, over a longer period, and ideally across more than one jurisdiction, since AML regulatory expectations vary a lot by country. It would also be worth tracking whether the model's performance holds up over time, or whether it needs regular retraining as laundering patterns shift, that's the kind of question three months of data can't really answer. What I think this study does establish is that the technology genuinely helps. Working out how to deploy it in a way regulators are comfortable with is really the next problem.

References

- Athari, S. A., Sassine, M. E., Agyemang, E. T., Kirikkaleli, D., & Saliba, C. (2026). Nonlinear Dynamics of AI Investment and Economic Growth in Germany: Evidence from a NARDL Approach. *Economies*, 14(1), 31.
- Azoury, N., Subrahmanyam, S., & Sarkis, N. (2024). The Influence of a Data-Driven Culture on Product Development and Organizational Success through the Use of Business Analytics. *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.*, 15(2), 123-134.
- Beainy, R., & Kamel, C. (2026). Gold and Bitcoin: A Quantitative Comparative Analysis. *The Banking and Finance Review*, 17(1).
- Bejjani, G., Toros, E., Elia, J., & Jabbour Al Maalouf, N. (2024). Technology analytic adeptness for examining the interrelatedness between the main macroeconomic variables in Lebanon employing a vector autoregression (VAR). *Journal of Decision Systems*, 1-30.
- Karam, R., & Fares, N. (2023). Machine learning approaches to anti-money laundering detection in Lebanese retail banking. *EuroMed Journal of Management*, 5(1), 55-71.
- Kaspard, J. A., Khalil, F. C., & Kamel, C. A. (2025). Factors contributing to non-financial performance of companies. In *Sustainable Economic Development: Utilizing Non-Financial Performance, Sustainability Reporting, and Corporate Governance* (pp. 413-442). Cham: Springer Nature Switzerland.
- Kaspard, J., Kamel, C., Khalil, F., & Beainy, R. (2026). Financial Performance, Risk, and Market Integration of Sustainability-Oriented Equity Indices: Implications for the Sustainability Transition (2010–2025). *Risks*, 14(5), 99.
- Khalife, D., Subrahmanyam, S., & Farah, A. (2024). A Sustainable Circular Business Model to Improve the Performance of Small and Medium-sized Enterprises Using Blockchain Technology. *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.*, 15(2), 240-250.
- Khalife, D., Yammine, J., El Bazi, T., Zaki, C., & Jabbour Al Maalouf, N. (2025). Analyzing the impact of investor sentiment on S&P 500 prices using deep learning models. *Journal of Financial Reporting and Accounting*, 23(4), 1365-1386.
- Saber, A. M., & Sassine, M. E. (2022). Reflections on internal control and internal audit roles for ensuring financial transparency toward Lebanese NGOs financial donors. *EuroMed Journal of Management*, 4(3), 186-198.
- Sawaya, C., Khalil, F. C., Jabbour Al Maalouf, N., & Kaspard, J. (2026). Revolutionizing audit practice in emerging economies: the role of Artificial Intelligence in enhancing the audit of consolidated financial statements. *Cogent Business & Management*, 13(1), 2625974.
- Subrahmanyam, S., Aishwaryalaxmi, N. S., Khalife, D., Shaikh, I. A. K., Faldu, R., & Asthana, N. (2024, April). Impact of knowledge management and big data analytics capabilities on firm performance. In *2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)* (pp. 1-5). IEEE.