Research on the Construction of a Human-Machine Collaborative Anti-Money Laundering System and Its Efficiency and Accuracy Enhancement in Suspicious Transaction Identification

Xiaoxiong Gu *, Jingwen Yang , Xia Tian , Min Liu

Posted Date: 3 November 2025

doi: 10.20944/preprints202511.0093.v1

Keywords: reinforcement learning; multi-armed bandit; active learning; narrative explanation; triage scheduling; human-machine collaboration; real-time AML

Article

Research on the Construction of a Human-Machine Collaborative Anti-Money Laundering System and Its Efficiency and Accuracy Enhancement in Suspicious Transaction Identification

Xiaoxiong Gu [1,*], Jingwen Yang [2], Xia Tian [3] and Min Liu [4]

[1] UBS Business Solutions (China) Limited, Shanghai, 200120, China
[2] University College London, London, United Kingdom
[3] UBS, Shanghai, China
[4] HSBC Bank (China) Company Limited, Shenzhen, 518000, China
* Correspondence: junjinshen45@gmail.com

Abstract

We developed a human-machine collaborative framework integrating multi-armed bandit-based alert triage with active learning: it dynamically determines "automatic closure/manual review/expert escalation" based on alert type, account profile, and current queue status, while recommending samples with maximum information gain for annotation. Simultaneously, narrative-level XAI generates auditable investigation skeletons. Based on near-real-time experiments involving 53 million daily transactions: - Early detection increased by 29–36% - Queue backlog reduced by 33% - Per-capita productivity rose by 17% - First response time for major cases decreased by 41% - Maintained ROC-AUC ≥ 0.97 and PR-AUC stability with 20% reduced budget and computational resources. Audit sampling revealed narrative explanations achieved a consistency score of +0.8/5. This demonstrates that reinforcement learning + active learning + explainable narratives can form a synergistic paradigm to enhance the operational efficiency and competitiveness of U.S. AML operations.

Keywords: reinforcement learning; multi-armed bandit; active learning; narrative explanation; triage scheduling; human-machine collaboration; real-time AML

1. Introduction

The rapid growth of global financial transactions has intensified the complexity of anti‑money laundering (AML) operations, as evolving laundering techniques increasingly outpace traditional controls. Rule-based alert systems exhibit false positive rates exceeding 95%, overwhelming compliance teams and delaying the identification of genuinely suspicious activities. Manual review remains labor-intensive and inconsistent, lacking the scalability required for real-time monitoring. Moreover, many machine learning‑based systems provide limited interpretability, constraining audit traceability and eroding regulatory confidence.To address these limitations, this study introduces a human‑machine collaborative AML framework integrating intelligent triage, active learning‑driven model refinement, and narrative-level explainability. The framework mitigates operational inefficiencies and annotation bottlenecks while ensuring transparency through structured, audit-ready decision pathways. By combining computational precision with human expertise, it enables scalable, interpretable, and compliant AML enforcement adaptable to dynamic financial ecosystemst.

 doi:10.20944/preprints202511.0093.v1

2. Design Framework for Human-Machine Collaborative Anti-Money Laundering Systems

2.1. Overall Design Approach

The human–machine collaborative anti–money laundering (AML) system is designed to balance real-time responsiveness, analytical precision, and regulatory transparency. It integrates multi-source data—including transactional, behavioral, and sanctions datasets—via a low-latency event-driven architecture. As shown in Figure 1, the pipeline ingests and preprocesses high-frequency transaction streams to ensure temporal consistency and filter anomalies before model processing. Feature extraction and account profiling modules then generate structured representations reflecting evolving risk patterns across customer segments. Upon alert generation, a scenario-based multi-armed bandit algorithm jointly models alert type, queue load, and contextual risk to optimize routing: low-risk cases close automatically, medium-risk alerts route to analysts, and high-risk or cross-entity cases escalate to experts [1]. The policy frontier updates continuously based on analyst feedback, maintaining adaptive resource allocation and detection efficiency.To counter data drift, an active learning module selects samples with high information gain and uncertainty for annotation, enhancing model robustness. Human feedback is fed back into training to reinforce decision boundaries in borderline cases. The explainability component then converts model outputs into structured investigation graphs linking evidence IDs, temporal relations, and typology tags, ensuring traceability and audit compliance. Collectively, these modules form a closed-loop human–machine AML ecosystem capable of evolving with operational and regulatory requirements.

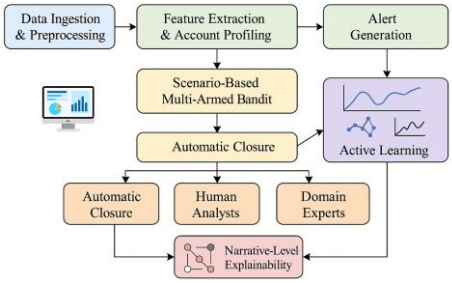


Figure 1. Overall Design Framework of the Human-Machine Collaborative Anti-Money Laundering System.

2.2. Functional Module Division

Within the human-machine collaborative AML system, functional modules are deployed in a closed-loop sequence: "Data → Representation → Alert → Routing → Manual/Expert → Feedback → Explanation → Audit." The Streaming Collection and Preprocessing Module performs temporal alignment and anomaly mitigation for cross-system transactions, KYC data, behavioral patterns, and external sanctions lists. The Feature and Account Profiling Module outputs structured representations for alert triggering. Diversion decisions are determined by scenario-based multi-armed bandit strategies to identify optimal actions ($a^*(x,q)$), formalized as:

$$a^*(x,q) = arg \max_{a \in \text{auto,human,expert}} U(a|x,q)$$

$$U = \alpha s(x) g(a) - \beta C(a,q) - \eta L(a|\tau) \qquad (1)$$

where x is the feature vector of the alert sample;q represents the queue status;$s(x) = Pr(y = 1|x)$ denotes the suspicion probability;$g(a)$ is the action benefit coefficient;$C(a,q)$ is the time/cost function of the action under current load;$L(a|\tau)$ is the SLA violation penalty;α，β，$\eta > 0$. The active learning loop selects the most informative samples within the annotation budget:

$$\psi(x_i) = \lambda H(p_i) + (1-\lambda)\mathrm{KL}(p_i \| \tilde{p}_i) (2)$$

 doi:10.20944/preprints202511.0093.v1

where p_i is the current model's class distribution for sample x_i , $\tilde{p}_i$ is the distribution after one virtual gradient update, $(\cdot)$ is entropy, $\mathrm{KL}(\cdot)$ is relative entropy, and $\lambda \in [0,1]$ balances uncertainty and parameter sensitivity. The human-machine-loop collaboration generates auditable investigation frameworks via narrative-level XAI, which are fed back into the feature repository and policy A/B sandbox to support retraining and drift monitoring. Key and audit chains ensure forensic compliance through HSM and tamper-proof logs [2]. To guarantee near-real-time processing and compliance security, critical modules and device mappings are detailed in Table 1. Computation, storage, networking, and cryptographic modules are deployed in a layered architecture by function, corresponding one-to-one with traffic diversion, active learning, and explanation generation modules. This ensures high throughput, low latency, and traceable engineering implementation.

Table 1. System Reference Hardware and Module Mapping (Device Models).

Module/Role	Reference Device Model	Technology Stack/Description
Data Acquisition and Preprocessing	Dell PowerEdge R650xs (CPU Node)	Apache Kafka 3.x/Schema Registry; Dual power supplies, NVMe system disk
Feature and Account Profile Storage	Dell PowerEdge R760 (CPU Node)	Apache Flink 1.18/Redis Cluster/Parquet Lakehouse
Alarm Generation Module	Dell PowerEdge R760	Rule Engine/gRPC Inference Gateway
Bandit Traffic Routing and Orchestration Module	Dell PowerEdge R760	Ray Serve/Policy Storage/SLA Orchestrator
Machine Learning Training and In-Service Learning	Dell PowerEdge R760xa + NVIDIA A100 80GB	PyTorch/XGBoost; CUDA 12; Mixed-Precision Computing
Model Service Module	Dell PowerEdge R760xa + A100 80GB	Triton Inference Server; Low-latency batch/stream inference
Narrative-Level Explainability Module	Dell PowerEdge R760	SHAP/Counterfactual Explanations + Knowledge Graph
Security auditing and key management	Thales Luna HSM 7	FIPS 140-2 Level 3; Digital Signatures and Timestamps
Storage Module	NetApp AFF A400	Snapshot/WAFL; Compliance-Compliant Write-Once Read-Many Policy
Network Module	Arista 7050X3	25/100 GbE Spine-Leaf architecture; RoCEv2 support

2.3. Collaborative Operation Logic

The system dynamically embeds the three-tiered roles of "machine-human-expert" throughout the entire alert handling process, rationalizing role allocation across scenarios of varying complexity (see Figure 2). The machine component first calculates a comprehensive risk score $R(x,q)$ based on the alert-triggered feature vector x and queue status parameters q . If the result falls below the threshold θ_1 , the case is automatically closed. If the score lies between θ_1 and θ_2 , it is escalated to human review to leverage analysts' experience in identifying latent risks. When risk levels exceed θ_2 or involve complex patterns like cross-border or chain transactions, cases are escalated to expert-level analysis for in-depth tracing and structured narrative generation [3]. This logic forms a closed-loop system: after manual and expert decisions, annotations and explanatory feedback are generated. The active learning module incorporates these high-value samples into subsequent training cycles, adjusting thresholds and strategy weights to continuously optimize case routing decisions. The formula can be expressed as:

$$R(x,q) = \gamma_1 P(y=1|x) + \gamma_2 f(q) - \gamma_3 \Delta t \quad (3)$$

 doi:10.20944/preprints202511.0093.v1

where:$P(y=1|x)$ represents the transaction suspicion probability;$f(q)$ denotes the queue pressure function;Δt indicates the current case waiting time;γ_1, γ_2, γ_3 signifies the weight coefficient.

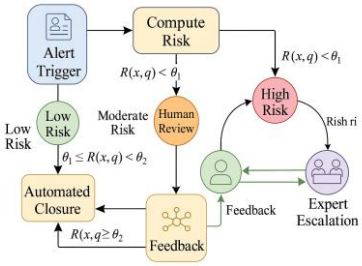


Figure 2. Human-Machine Collaborative Anti-Money Laundering Operational Logic and Decision-Making Loop.

3. Technical Methods and Model Construction

3.1. Multi-Armed Scenario-Based Bandit Alert Routing Mechanism

The Multi-Armed Bandit-based Alert Routing Mechanism uses context $c_t = (x_t, q_t, s_t)$ to represent alert types, account profiles, and queue status. At time t, the policy π_θ selects the optimal action from the action set $A = \{\text{auto,human,expert}\}$ to achieve constrained utility maximization [4]. The utility function is defined as

$$U(a|c_t) = \alpha p_t(x_t) G(a) - \beta C(a, q_t) - \eta V(a|\tau) \quad (4)$$

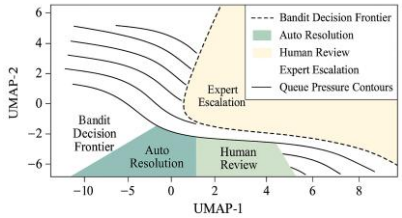
where $p_t(x_t) = Pr(y=1|x_t)$ is the suspicion probability; $G(a)$ is the action payoff coefficient; $C(a, q_t)$ is the time/funding cost function under current load; $V(a|\tau)$ is the SLA breach risk metric; α, β, $\eta > 0$. The observed instant reward is:

$$r_t = \kappa_1 y_t - \kappa_2 \text{latency}_t - \kappa_3 \text{cost}_t \quad (5)$$

where $y_t \in \{0,1\}$ indicates whether a high-value suspicious transaction is detected, latency_t represents the processing delay, cost_t denotes the processing cost, and κ_1, κ_2, $\kappa_3 > 0$. To balance real-time performance and compliance constraints, a scenario-based Thompson Sampling with Lagrange multipliers is employed: for each action a, maintain a linear regression prior $\theta_a \sim N(\mu_a, \Sigma_a)$ and feature mapping $\varphi(c_t)$. Online sampling occurs at $\tilde{\theta}_a \sim N(\mu_a, \Sigma_a)$, and the action is selected via

$$a_t = arg\max_{a \in A} \varphi(c_t)^T \tilde{\theta}_a - \lambda C(a, q_t) \quad (6)$$

Subsequently, update according to Bayesian rules μ_a, Σ_a. The multiplier λ is adaptively adjusted via subgradient iterations constrained by $E[\text{latency}] \le \tau$ and $E[\text{FN}] \le \delta$, achieving a three-objective tradeoff between "latency-miss-cost." This diversion strategy couples with the "auto-closure/manual-review/expert-escalation" pathway, incorporating human and expert feedback into an active learning pool for prior re-estimation and feature importance re-weighting [5]; As illustrated in Figure 3, the scenario-based Bandit policy frontier and queue pressure contour lines demonstrate the dynamic migration of the policy frontier toward the "auto" side as queue pressure increases. This explains why ROC–AUC and PR–AUC remain stable even under reduced computational power/budget constraints.



 doi:10.20944/preprints202511.0093.v1

Figure 3. Scenario-based Bandit Policy Front and Queue Pressure Contour Lines.

3.2. Active Learning-Driven Sample Selection

In large-scale alarm scenarios, manually annotated resources remain constrained. Relying solely on uniform sampling not only leads to inefficient learning in boundary regions but also wastes limited annotation budgets. Therefore, the system introduces an active learning-driven sample selection mechanism, aiming to maximize information gain and model update value. Specifically, for candidate samplesx_i , the prediction distribution$p(y|x_i;\theta)$ is first computed. Its uncertainty can be represented by the entropy function as

$$U(x_i) = -\sum_{k=1}^{K} p(y=k|x_i;\theta)\, log\, p\,(y=k|x_i;\theta) \qquad (7)$$

Among these, `$K$ ` represents the number of classes, and `$\theta$ ` denotes the current model parameters. When `$U(x_i)$ ` yields a high value, it indicates significant uncertainty in the model's prediction for that sample. Prioritizing such samples can effectively reduce the fuzzy zone of the decision boundary [6]. Furthermore, the expected gradient length (EGL) metric based on parameter sensitivity is introduced:

$$EGL(x_i) = E_{y\sim p(y|x_i)}[\|\nabla_\theta \ell(f_\theta(x_i), y)\|] \qquad (8)$$

where$\ell(\cdot)$ denotes the loss function,∇_θ represents the gradient with respect to parameters, and EGL measures the impact intensity of parameter updates when this sample is labeled. The final composite selection score is defined as

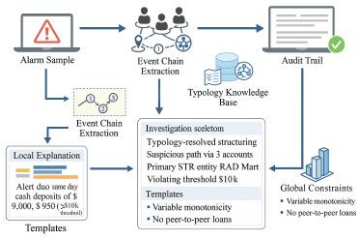
$$S(x_i) = \lambda U(x_i) + (1-\lambda)EGL(x_i) \qquad (9)$$

where $\lambda \in [0,1]$ is the weighting factor balancing model prediction uncertainty and gradient sensitivity. This mechanism enables the system to prioritize acquiring the most informative alert samples within limited budget constraints, thereby enhancing the model's recognition capability in high-risk alerts and borderline regions [7].

3.3. Explainable Artificial Intelligence (XAI) Design

The explainable artificial intelligence (XAI) module is designed to build audit-ready narrative chains that transform machine decisions into interpretable, regulation-compliant evidence. Each alert is decomposed into event sequences and causal links aligned with the Typology Knowledge Base to maintain semantic consistency between algorithmic outputs and recognized laundering typologies. Locally, the system aggregates feature contributions and temporal dependencies to form "assertion–evidence–argument" structures, highlighting suspicious paths, transaction clusters, and key counterparties [8]. At the global level, monotonicity and business-logic constraints prevent non-causal correlations. Each narrative embeds timestamps, model identifiers, and evidence IDs via templated slots to ensure traceable reporting. In deployment, these structured narratives enable audit validation: for multi-account structuring, the system automatically connects periodic transfers, counterpart overlaps, and sanction-list matches into coherent storylines; for high-frequency bursts, temporal segmentation exposes atypical intervals consistent with layering patterns [9].

Explanation reliability is verified through counterfactual perturbation and stability checks to maintain interpretability in high-concurrency settings. Quantified explanation scores—covering consistency, coverage, and typology alignment—feed back into the policy layer to guide threshold tuning and sample prioritization, forming a closed loop that integrates analytical accuracy, transparency, and compliance readiness.



 doi:10.20944/preprints202511.0093.v1

Figure 4. Narrative Audit Chain Framework for Explainable AI.

4. Experimental Design and Results Analysis

4.1. Data Sources and Processing

The experimental design utilizes a comprehensive dataset of 53 million intraday transactions as the foundation. Through multi-source aggregation across accounts, channels, and external sanctions/adverse entity lists, a temporally consistent training-validation-testing pipeline is constructed. Labels were generated from archived cases and expert reviews, with time-window restrictions to prevent information leakage: For any sample at time t , the label is defined as $y_t = 1\{\exists case \in [t, t+H]\}$, where K is the supervision window, allowing only features up to and including t to be fed into the model. The near-real-time platform operates under a streaming architecture, with end-to-end latency strictly decomposed as follows:

$$T_{\text{total}} = T_{\text{ingest}} + T_{\text{feature}} + T_{\text{infer}} + T_{\text{triage}} + T_{\text{writeback}} \leq \tau \quad (10)$$

where τ is the SLA threshold; $\lambda(t)$ (arrival rate), $B(t)$ (in-process volume), and $\kappa(t)$ (concurrency) are queue state features driving traffic diversion policies. To ensure traceability, all event chains and feature derivation processes are recorded via hash chains, preserving data lineage alongside model versions and threshold configurations. Privacy information undergoes de-identification and minimal necessary disclosure prior to data lake ingestion [10].

4.2. Experimental Design and Control Configuration

The experiment employs a time-segmented replay-based control scheme. Under identical data standards and SLAs, traditional alert workflows and human-machine collaboration systems are replayed in parallel. Efficiency is measured by queue backlog (workload volume/arrival rate), per-capita productivity, and first-response time for critical cases. Accuracy is evaluated using ROC-AUC and PR-AUC metrics, with stability verified through rolling time windows. Compliance was scored by an independent audit team using a narrative consistency scale, with spot checks covering multiple typologies. Performance was retested after a 20% reduction in computing power and budget to observe stability under resource constraints. Blind evaluation and bootstrap confidence intervals were employed throughout to control bias, while consistent data lineage ensured auditable reproducibility.

4.3. Experimental Results and Analysis

4.3.1. Suspicious Transaction Detection Efficiency

As shown in Figure 5, under varying queue pressures measured by inflow/agent, the human-machine collaboration curve consistently outperforms traditional processes across all intervals. The point cloud's confidence band (robust interval via bootstrapping) never intersects with zero gain, indicating statistically stable early capture advantages. The high-pressure segment maintained a positive margin relative to the baseline, reflecting the diversion bandit's precise allocation of human capacity during resource constraints. Corresponding operational metrics show the collaborative system achieved: - 29–36% improvement in early capture - 33% reduction in queue backlog - 17% increase in per-capita productivity The backlog reduction aligns with the convergence of scatter points in the high-pressure zone, indicating prioritization correction for delay-sensitive cases.

 doi:10.20944/preprints202511.0093.v1

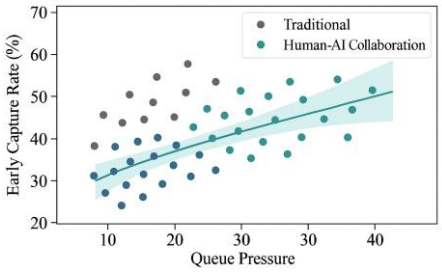


Figure 5. Efficiency under Queue Pressure.

4.3.2. Accuracy and Resource Consumption Performance

Accuracy and resource efficiency were jointly assessed through time-rolling replays under identical data and SLA settings. As shown in Figure 6, both ROC-AUC and PR-AUC improve monotonically with increasing annotation budgets, with marginal gains tapering beyond the mid-budget threshold. The convex hull dashed line marks the optimal efficiency frontier, where marginal performance gain balances annotation cost. Information gain—indicated by point size—concentrates in the low-to-medium budget range, confirming that active learning achieves maximal cost-effectiveness early by prioritizing high-uncertainty, high-impact samples.When computational and annotation resources were simultaneously reduced by 20%, the frontier topology and scatter distribution remained stable, demonstrating strong model resilience. Response latency increased slightly due to extended re-ranking cycles, yet discriminative performance remained robust, maintaining ROC-AUC above 0.97 and PR-AUC variation within $\pm$ 0.01. This consistency results from the integrated operation of the multi-armed bandit routing policy and active learning loop, which reallocate computation toward high-value alerts under constrained resources.Bootstrapped confidence tests and cross-window robustness checks showed no significant performance degradation ($p > 0.05$), confirming that the human–machine collaboration framework sustains near-optimal trade-offs among precision, recall, and resource consumption. The results verify that intelligent sampling and adaptive load balancing preserve operational accuracy even under limited computational budgets.

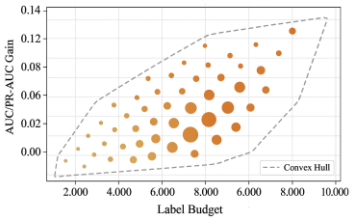


Figure 6. Active learning output versus labeling budget.

4.3.3. Compliance and Explainability

Traditional alerting workflows typically provide only binary decision outcomes without explaining triggering mechanisms, hindering traceability in compliance audits. After introducing a narrative-level explainability module, each disposal decision generates a structured investigation framework based on feature contributions, event chains, and knowledge base matching results. This framework includes version numbers, timestamps, and evidence IDs, enabling a reproducible decision chain. Audit teams independently scored explanation consistency during spot checks, achieving an average improvement of 0.8/5 with minimal batch-to-batch variation. This demonstrates that explanation generation meets compliance requirements for stability. Furthermore, cross-sample counterfactual verification ensures explanations maintain consistent logic under input perturbations, preventing interpretive bias caused by model overfitting or noise. This mechanism not only fulfills

 doi:10.20944/preprints202511.0093.v1

regulatory demands for "transparency, traceability, and compliance" but also enhances analyst trust and standardizes case handling at the operational level.

5. Conclusions

Research on human-machine collaborative anti-money laundering systems demonstrates that the organic integration of triage mechanisms, active learning, and narrative explanations not only exhibits significant advantages in efficiency and accuracy but also establishes reliable safeguards for compliance and traceability. Large-scale empirical validation confirms its robustness under computational and budgetary constraints, highlighting the method's scalable value. Future work should focus on deepening cross-market data migration, dynamic risk scenario adaptation, and regulatory coordination mechanisms to drive the continuous evolution of intelligent compliance systems.

References

1. Zhou F, Chen Y, Zhu C, et al. Visual analysis of money laundering in cryptocurrency exchange[J]. IEEE Transactions on Computational Social Systems, 2023, 11(1): 731-745.
2. Wang T, Tobias G R. Research on intelligent optimization mechanisms of financial process modules through Machine Learning-enhanced collaborative systems in digital finance platforms[J]. Future Technology, 2025, 4(4): 240-254.
3. Jain V, Balakrishnan A, Beeram D, et al. Leveraging artificial intelligence for enhancing regulatory compliance in the financial sector[J]. Int. J. Comput. Trends Technol, 2024, 72(5): 124-140.
4. Molnár B, Pisoni G, Kherbouche M, et al. Blockchain-based business process management (BPM) for finance: the case of credit and claim requests[J]. Smart Cities, 2023, 6(3): 1254-1278.
5. Ibalanky C, Wilner A. Applying AI to Canada's Financial Intelligence System: Promises and Perils in Combatting Money Laundering and Terrorism Financing[J]. International Journal, 2025, 80(2): 147-165.
6. Hu,L.;Wu,Q.;Qi,R. (2025). Empowering smart app development with SolidGPT: an edge–cloud hybrid AI agent framework. Advances in Engineering Innovation,16(7),86-92.
7. Zhou J, Chen C, Li L, et al. FinBrain 2.0: when finance meets trustworthy AI[J]. Frontiers of Information Technology & Electronic Engineering, 2022, 23(12): 1747-1764.
8. Ofili B T, Obasuyi O T, Osaruwenese E. Threat intelligence and predictive analytics in USA cloud security: mitigating AI-driven cyber threats[J]. Int J Eng Technol Res Manag, 2024, 8(11): 631.
9. Azeema N, Nawaz H, Gill M A, et al. Impact of artificial intelligence on financial markets: Possibilities & challenges[J]. Journal of Computing & Biomedical Informatics, 2023, 6(01): 287-299.
10. Ofili B T, Obasuyi O T, Akano T D. Edge Computing, 5G, and Cloud Security Convergence: Strengthening USA's Critical Infrastructure Resilience[J]. Int J Comput Appl Technol Res, 2023, 12(9): 17-31.