European Journal of AI, Computing & Informatics

Vol. 1 No. 3 2025

Article **Open Access**

Leveraging Large Language Models for Anomaly Event Early Warning in Financial Systems

Luqing Ren [1,*]

[1] Columbia University, New York, NY, USA
* Correspondence: Luqing Ren, Columbia University, New York, NY, USA

Received: 19 September 2025
Revised: 25 September 2025
Accepted: 09 October 2025
Published: 16 October 2025

Abstract: Given the strong semantic concealment, complex data structures, and high responsiveness requirements of anomaly events in financial systems, this study investigates advanced anomaly early warning methods that integrate large language models (LLMs) to enhance detection and response capabilities. The research systematically outlines semantic processing approaches for multi-source heterogeneous data, including structured transaction records, unstructured textual reports, and real-time market indicators, to ensure comprehensive analysis of potential anomalies. It further details the construction path for risk identification models, emphasizing feature extraction, multi-dimensional correlation analysis, and adaptive learning mechanisms that allow the system to dynamically adjust to evolving financial patterns. In addition, the study presents the generation and push mechanism for early warning information, ensuring that alerts are delivered promptly to relevant stakeholders while maintaining high interpretability and actionable insight. The system's response efficiency and recognition performance are rigorously evaluated in real trading environments, demonstrating the framework's robustness under high-frequency and high-volume transaction conditions. Comparative experiments with various baseline models show that the proposed approach exhibits strong adaptability and practical value, achieving superior performance in both recognition accuracy and response timeliness. These results indicate the potential of LLM-enhanced anomaly early warning systems to support more reliable and intelligent risk management in complex financial environments.

Keywords: financial anomaly events; large language models; multi-source data fusion; early warning systems

1. Introduction

Frequent abnormal events in financial systems exhibit characteristics such as complex data types, concealed behavioral chains, and stringent response timeliness requirements, rendering traditional rules and shallow model's incapable of achieving precise identification and dynamic early warning [1]. To enhance the perception of potential risks in unstructured text and multimodal behavioral data, this study constructs an anomaly modeling and early warning framework centered on large language models. By systematically integrating semantic extraction, contextual analysis, and risk grading mechanisms, the framework covers critical stages including transaction monitoring, warning generation, and real-time push notifications. It aims to improve anomaly recognition accuracy and response loop capabilities within financial systems, providing technical support for intelligent risk control systems.

2. Risk Characteristics of Anomalous Events in Financial Systems

Anomalous events within financial systems exhibit high levels of concealment, suddenness, and cross-domain propagation. They primarily encompass types such as irregularity unlet transactions, illegal fund transfers, account compromise, and abnormal volatility [2]. Such incidents often deviate from normal business patterns, exhibiting distinct data anomalies-such as frequent small-value transfers, surges in off-hours transactions, or circular fund flows between specific accounts-and frequently utilize unstructured textual information (e.g., reports, announcements, public sentiment) as implicit signals. The complexity of these risks stems not only from diverse business structures and heterogeneous data but also from the temporal interdependencies and contextual dependencies inherent in event evolution. This makes accurate identification and timely alerts challenging for traditional rule-based or statistical methods. Furthermore, masking strategies for anomalous behavior continuously evolve-such as employing multi-hop accounts and mixed transaction paths to disrupt tracking-increasing modeling complexity [3]. Consequently, a technical framework with deep semantic understanding and long-range dependency modeling capabilities must be introduced to provide foundational risk feature support for subsequent development of highly robust large language models.

3. Modeling Methods for Financial Anomalies Using Large Language Models

3.1. Semantic Understanding Capabilities of Large Language Models

Leveraging context modeling capabilities acquired through large-scale pre-training, large language models can effectively capture hidden anomalous semantic patterns within financial contexts [4]. Traditional methods often struggle to handle contextual dependencies and semantic ambiguities in unstructured text during anomaly event modeling. In contrast, large language models can achieve high-precision modeling of key entities, transaction motivations, and event chains within lengthy texts through self-attention mechanisms. For instance, when identifying announcements suspected of financial irregularities or insider trading, the model can not only comprehend explicit descriptions but also extract risk signals embedded within the context (such as phrases like "short-term concentrated share accumulation" or "undisclosed funding sources"). Furthermore, LLMs possess semantic transfer capabilities, enabling them to generalize abstract semantics across different financial sub-scenarios (e.g., banking operations, securities announcements, online sentiment analysis) to support unified modeling. Their robust semantic representation capabilities lay the foundation for subsequent text encoding, risk scoring, and multimodal information fusion, significantly enhancing the perception of potential abnormal events and judgment accuracy.

3.2. Representation and Processing of Anomaly Event Text Data

Financial anomalies often manifest through unstructured texts like announcements, sentiment data, and transaction notes. These sources present fragmented information with implicit meanings, requiring meticulous processing to support downstream modeling. First, the system performs noise cleaning and word segmentation on raw text, leveraging domain-specific lexicon to identify professional entities (e.g., account names, fund flows, risk control terminology). Subsequently, context-enhanced entity annotation strategies are employed to construct financial semantic units, leveraging pre-trained language models to generate vector representations. A typical approach involves sentence vector expressions based on token embeddings:

$$V_{doc} = \frac{1}{n}\sum_{i=1}^{n} V_{w_i} \tag{1}$$

Where: V_{w_i} denotes the word vector for the$i$ th word, and V_{doc} represents the semantic vector for the entire text segment. In practice, to enhance anomaly representation, a word weighting mechanism can be introduced:

$$V_{doc} = \sum_{i=1}^{n} \alpha_i \cdot V_{w_i}, where\ \alpha_i = \frac{exp(s_i)}{\sum_{j=1}^{n} exp(s_j)} \tag{2}$$

where: α_i is the importance weight of the i th word, and s_i is computed via an attention mechanism, reflecting the semantic strength of high-risk vocabulary in financial statements. The resulting high-dimensional vector input to subsequent models preserves event chronology, semantic logic, and risk implications, providing stable semantic support for anomaly detection.

3.3. Model Training and Parameter Optimization Strategy

To address the challenges of high-dimensional sparse inputs and extreme class imbalance in financial anomaly detection, the model training phase requires optimization strategies that balance gradient stability and generalization capability. A cross-entropy-based loss function is adopted to enhance recognition accuracy for high-risk samples. The specific form is as follows:

$$L = -\sum_{i=1}^{N} y_i \cdot log(\hat{y}_i) + (1 - y_i) \cdot log(1 - \hat{y}_i) \quad (3)$$

Where: $\hat{y}_i$ represents the model prediction probability, and y_i denotes the label. To adapt to the distribution characteristics of different financial subtasks, learning rate scheduling and momentum adjustment strategies are introduced during training, with the AdamW optimizer employed to control overfitting.

For parameter optimization, grid search combined with cross-validation was employed to conduct combinatorial experiments on key hyperparameters such as learning rate, batch size, and layer freezing strategy. Table 1 presents model performance under selected configurations, revealing significant variations in accuracy, F1 score, and training stability across different settings. Table 1 indicates that moderately freezing model-level semantic weights, combined with mini-batch training and low learning rate settings, enhance generalization performance and stability in anomaly detection. This provides data support for subsequent strategy adjustments during system deployment.

Table 1. Model performance under different hyperparameter settings.

Learning Rate	Batch Size	Optimizer	Frozen Layers	Accuracy (%)	F1 Score	Epochs
1e-05	16	AdamW	Top 6	87.43	0.852	12
2e-05	32	AdamW	None	84.91	0.823	10
3e-05	16	SGD	Top 3	81.26	0.792	15
1e-05	64	AdamW	All Frozen	78.03	0.751	20

4. Design of the Early Warning Framework for Anomalous Events

4.1. Overall Architecture of the Early Warning System

To achieve proactive identification and real-time response to abnormal events in financial systems, the early warning system employs a modular architecture comprising data acquisition, fusion computation, model inference, and result feedback. The overall architecture is illustrated in Figure 1. The system adopts a layered decoupled design to ensure efficient and stable processing workflows. At the data ingestion layer, the system supports parallel access to multi-source data, including structured transaction logs, unstructured announcement texts, and external sentiment data, forming a dynamic data stream. The processing layer performs data cleansing, label completion, semantic enhancement, and feature unified encoding on ingested data while simultaneously generating risk alert signal streams. The inference layer deploys large language models and fusion anomaly detection algorithms to execute context-based semantic event judgment and scoring, outputting risk level labels. Finally, the alert output layer pushes anomaly events to financial regulators, risk control platforms, or business terminals, forming a closed-loop response pathway. Figure 1 illustrates the collaborative relationships among modules within this system architecture, providing a foundational structure for subsequent modules such as feature fusion and risk grading.

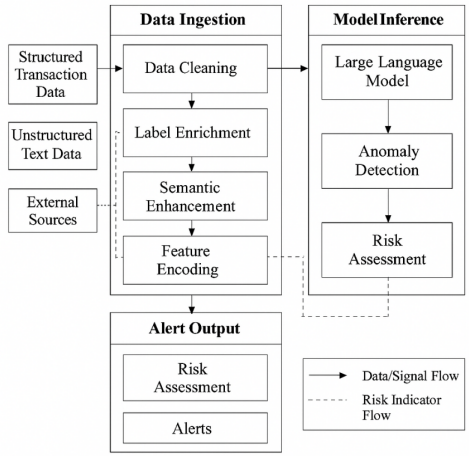

Figure 1. Architecture Diagram of the Financial Anomaly Early Warning System.

4.2. Multi-Source Data Fusion and Feature Extraction

Financial early warning tasks face challenges such as the coexistence of structured and unstructured data and the widespread distribution of multi-source heterogeneous information. To achieve a unified representation for model inputs, the system constructs a multi-source data fusion mechanism based on semantic alignment and temporal alignment. Structured data-such as account transaction logs, amount changes, and login behaviors-is directly mapped into quantitative features. While unstructured data like announcement texts, user interaction logs, and social sentiment are processed through large language models to extract embedding vectors, capturing deep semantic risk signals. After feature engineering, diverse data types form the subspace feature vector sets: F_s, F_t, and F_u, representing structured, textual, and user behavior dimensions respectively. The system employs a weighted concatenation strategy for unified representation:

$$F_{fusion} = \alpha \cdot F_s + \beta\, F_t + \gamma \cdot F_u \quad (4)$$

Where: α, β, γ represents the fusion weights obtained through adaptive learning based on feature importance. This approach preserves the uniqueness of each modality's data while ensuring overall semantic consistency and model acceptability.

To further enhance the model's discriminative capability, time-series windowing processing and risk context summarization mechanisms are introduced. This ensures feature inputs accurately reflect the dynamic evolution trajectory of anomalous events. The fused F_{fusion} serves as direct input for subsequent anomaly detection and risk grading modules, supporting highly robust judgments.

4.3. Anomaly Detection and Risk Grading Mechanism

Following multi-source feature fusion, the system performs high-precision detection of potential anomalies and constructs actionable risk grading mechanisms. Integrating semantic representations and contextual features generated by large language models, the system employs an anomaly detection algorithm based on confidence scoring. The core approach evaluates the semantic deviation between input samples and the training distribution [5], forming an anomaly scoring function:

$$A(x) = 1 - P_\theta(y|x) \quad (5)$$

Where: $P_\theta(y|x)$ represents the model's predicted probability for the input sample x based on current parameters θ. A lower value indicates greater abnormality. By modeling high-frequency behavioral patterns in financial tasks, the system can classify inputs with significantly deviant scores as high-risk events.

For risk classification, the system employs a three-tier risk grading strategy defined as “High Risk (R3)”, “Medium Risk (R2)”, and “Suspicious Monitoring (R1)”, utilizing a

threshold model constructed with expert annotations from historical financial events. Using real financial operations as an example: - If the anomaly score is $A(x) > 0.85$ and the event involves cross-account transfers or hidden path jumps, it is classified as R3, triggering an alert push – If $0.65 < A(x) \leq 0.85$, it is classified as R2, retaining the system alert without prompting the terminal – All others are classified as R1 and placed under monitoring.

Additionally, the system incorporates behavioral tag intervention strategies. By dynamically adjusting risk assessments based on supplementary dimensions like transaction timing, login IP, and session trajectories, it ensures high adaptability across diverse scenarios. This scoring and classification logic provides structured input to the alert output module, enhancing the interpretability and controllability of event identification.

5. System Implementation and Application Scenarios

5.1. Financial Transaction Monitoring Scenario Design

Within financial systems, the transaction layer is a high-risk zone for anomalies, encompassing multiple risk dimensions such as abnormal fund transfers, suspicious account activities, and disguised transaction paths. To effectively monitor potential anomalies and enable real-time alerts, the system implements scenario-based transaction monitoring centered around core transaction pathways. First, during data ingestion, the system interfaces with core payment clearing platforms, third-party payment gateways, online banking systems, and ATM channels to collect key fields such as transaction logs, account statuses, and terminal information in real time. Daily transaction monitoring capacity reaches up to 300 million records. The collection layer utilizes Kafka message queues and ETL pipelines to achieve high-throughput, low-latency data synchronization, ensuring millisecond-level event detection. Second, at the feature extraction and model embedding layer, the system constructs unified feature encoding based on transaction dimensions (transaction type, time, amount, account pair). This is combined with historical behavior vectors, geographic location deviations, and behavior chains from related accounts for modeling. Within the processing chain, large language models are embedded in the pre-transaction risk scoring phase. They perform semantic recognition and risk scoring on text-based fields-such as transaction remarks and counterparty account registration details-enhancing detection of disguised anomalies.

To accelerate response times, the system implements five real-time detection nodes during transaction monitoring: entry validation, transaction flow monitoring, session behavior modeling, model scoring, and threshold decision-making. Each node incorporates local caching and re-scoring mechanisms, enabling microsecond-level blocking at the earliest signs of anomalies to prevent risk propagation. The scenario design also comprehensively considers the evolutionary patterns of abnormal behavior chains, incorporating sliding window mechanisms and cross-account graph modeling methods to continuously track suspicious transaction chains. Through the deployment of this transaction monitoring design, the system gains the ability to perceive and trigger responses to typical abnormal patterns such as “multi-hop financial irregularities” “decentralized transfer and reconvergence” and “scheduled automated transactions” providing stable inputs for subsequent alert generation and risk control handling.

5.2. Alert Generation and Push Mechanism

After the transaction monitoring module identifies potential anomalies, the system converts these findings into structured alert information and executes multi-dimensional push notifications to achieve closed-loop management encompassing risk control intervention, operational coordination, and regulatory synchronization. Alert generation relies on risk grading results from front-end models, contextual tags, and behavioral chain modeling feedback. The system constructs standardized alert objects per event, comprising seven fields: transaction ID, account identifier, risk level, trigger factor, confidence range,

trigger time, and visual path summary-ensuring complete traceability. Alert thresholds are dynamically configured per scenario. For instance, in payment channel scenarios, transactions with confidence below 0.15 that simultaneously exhibit terminal geographic drift or associations with historically flagged accounts automatically generate high-priority R3 alerts. Approximately 21,000 structured alert records are generated daily, asynchronously distributed via embedded message brokers (e.g., Kafka, RocketMQ) to ensure downstream processing efficiency.

The push mechanism employs tiered routing: R3 events are pushed in real-time to manual risk control terminals and trigger temporary account freeze instructions; R2 events are included in the next day's risk control review list; R1 events are stored in the alert repository for dynamic learning by strategy models. The system also supports integration with CRM platforms to send risk warning SMS or app notifications to high-frequency abnormal users, enhancing customer response efficiency. Additionally, the alert design incorporates model explanation fields (e.g., ranking of primary trigger factors, historical behavior comparison summaries) to enhance the response efficiency and judgment accuracy of terminal operators while providing technical transparency for regulatory oversight.

6. Experiments and Effect Validation

6.1. Experimental Dataset and Evaluation Metrics

To evaluate the effectiveness of the financial anomaly alert system, the experiment utilized transaction data and business log samples from a joint-stock commercial bank covering Q2 to Q3 2023. This dataset includes structured transaction records, unstructured transaction notes, account operation logs, and blacklist tag information. The original dataset comprised 120 million records. After cleaning and filtering, a training and testing set of 2.65 million high-quality samples was constructed, with approximately 2.8% labeled as anomaly events. Unstructured text sources primarily included transaction descriptions from the mobile app, customer service records, and public sentiment reports, which were fused with structured features after BERT encoding.

Regarding evaluation metrics, the system adopts Precision, Recall, F1-score, and AUC as primary indicators to reflect identification accuracy and coverage, addressing the dual requirements of sensitivity and responsiveness in financial risk control. Response latency and recall time window length are introduced as system-level assessment parameters to ensure real-time usability during actual deployment. Furthermore, to mitigate label imbalance effects, experiments employ weighted F1 and grouped AUC analysis strategies to enhance metric stability and result interpretability.

6.2. Early Warning Model Performance Comparison

To validate the adaptability and performance advantages of large language models in identifying financial anomalies, the system compares four models-XGBoost, LightGBM, FinBERT, and the optimized LLM-FinRisk-using the same dataset and metric framework. Results are presented in Table 2. LLM-FinRisk demonstrated the most outstanding performance in Recall (0.890) and F1-score (0.861), reflecting its superior capability in identifying complex, high-potential-risk events compared to other models. Although XGBoost and LightGBM exhibit fast inference speeds, they show significant shortcomings in identifying low-frequency risks, with Recall values below 0.71. FinBERT achieved an excellent AUC (0.901) but had a slightly lower overall F1 score due to the lack of a structured feature integration mechanism. Regarding inference efficiency, LLM-FinRisk controlled latency through model pruning and hierarchical inference, achieving an average response time of 19.8ms, meeting the real-time requirements for online trading scenarios. Comprehensive analysis indicates that incorporating multimodal semantic modeling and risk context fusion mechanisms enhances LLM's generalization and practical applicability, providing robust support for future early warning system upgrades.

Table 2. Performance comparison of different anomaly detection models.

Model	Precision	Recall	F1-score	AUC	Inference Time (ms)
XGBoost	0.842	0.683	0.754	0.869	12.4
LightGBM	0.831	0.701	0.76	0.874	10.7
FinBERT	0.799	0.763	0.78	0.901	46.2
LLM-FinRisk	0.835	0.89	0.861	0.912	19.8

6.3. System Response Efficiency Analysis

To evaluate the real-time processing capability of the early warning system in high-frequency financial transactions, system response efficiency tests were conducted on typical data channels during peak transaction periods, covering three scenarios: mobile payments, online banking transfers, and ATM channels. The tests primarily focused on event ingestion latency, model distinction delay, early warning generation time, and overall end-to-end response time. Results are shown in Table 3. In the mobile payment scenario, the system maintained an average response time under 62 ms, meeting the requirement for sub-second response. In the ATM channel, due to complex device access links, the maximum latency increased to 95 ms. The system demonstrated stable performance under high concurrency (5000 TPS), with latency fluctuations below 8 ms, indicating robust throughput control and cache scheduling capabilities. Concurrently, the alert generation module's latency contribution was maintained below 22%, reflecting high compatibility between model embedding and message queue mechanisms, thereby providing a solid foundation for multi-channel risk perception.

Table 3. System latency metrics under different transaction scenarios.

Scenario	Avg Latency (ms)	Max Latency (ms)	Jitter (ms)	Max Throughput (TPS)
Mobile Payment	61.7	74.2	5.4	5800
Online Banking	68.3	82.6	6.2	5100
ATM Channel	72.4	95.1	7.9	4600

7. Conclusion

For highly concealed, context-dependent anomalies within financial systems, we have developed an early warning framework integrating multi-source semantic modeling with real-time risk perception. This demonstrates the technical adaptability and scenario generalization capabilities of large language models in extracting unstructured information and identifying abnormal behaviors. Future work may further expand model transfer and causal reasoning capabilities across markets and multilingual environments, enhancing the system's ability to holistically model and interpret complex financial irregularity chains.

References

1. Y. Nie, Y. Kong, X. Dong, J. M. Mulvey, H. V. Poor, Q. Wen, and S. Zohren, "A survey of large language models for financial applications: Progress, prospects and challenges," *arXiv preprint arXiv:2406.11903*, 2024.
2. Y. Liu, N. Bu, Z. Li, Y. Zhang, and Z. Zhao, "AT-FinGPT: Financial risk prediction via an audio-text large language model," *Finance Research Letters*, vol. 77, p. 106967, 2025. doi: 10.1016/j.frl.2025.106967
3. J. Lee, N. Stevens, and S. C. Han, "Large language models in finance (finllms)," *Neural Computing and Applications*, pp. 1-15, 2025. doi: 10.1007/s00521-024-10495-6
4. J. Wang, J. Liu, J. Pu, Q. Yang, Z. Miao, J. Gao, and Y. Song, "An anomaly prediction framework for financial IT systems using hybrid machine learning methods," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 11, pp. 15277-15286, 2023.

5. E. E. Alharasis, H. Haddad, M. Shehadeh, and A. S. Tarawneh, "Abnormal monitoring costs charged for auditing fair value model: evidence from Jordanian finance industry," *Sustainability*, vol. 14, no. 6, p. 3476, 2022. doi: 10.3390/su14063476
6. L. Ren, "Causal Inference-Driven Intelligent Credit Risk Assessment Model: Cross-Domain Applications from Financial Markets to Health Insurance," *Academic Journal of Computing & Information Science*, vol. 8, no. 8, pp. 8-14, 2025.
7. L. Ren, "Boosting Algorithm Optimization Technology for Ensemble Learning in Small Sample Fraud Detection," *Academic Journal of Engineering and Technology Science*, vol. 8, no. 4, pp. 53-60.