The American Journal of Interdisciplinary Innovations and Research
ISSN 2642-7478 Volume 08 - 2026

Explainable Behavioral Analytics for Financial Crime Detection Using Large Language Models

Aswin Sivaselvan

Independent Researcher, USA

Received: 26 Feb 2026 | Received Revised Version: 16 Mar 2026 | Accepted: 16 Apr 2026 | Published: 30Apr 2026

Volume 08 Issue 04 2026 |

Abstract

Financial crime continues to evolve through increasingly sophisticated behavioral patterns that are difficult to detect using conventional rule-based monitoring systems. Advances in large language models (LLMs) have created new opportunities for analyzing unstructured textual information, behavioral indicators, and contextual evidence associated with suspicious financial activities. This paper proposes an Explainable Behavioral Analytics framework that combines large language models with machine learning techniques to improve financial crime detection while ensuring transparency and interpretability. The framework integrates structured transaction data with unstructured textual information, including investigation reports, customer interactions, suspicious activity narratives, and compliance documentation. Behavioral features extracted using transformer-based language models are combined with supervised classification algorithms to identify potentially fraudulent activities. Explainability mechanisms are incorporated to provide investigators with understandable reasoning behind model predictions, thereby improving trust, regulatory compliance, and operational decision-making. Experimental analysis using representative financial crime datasets demonstrates that the proposed framework enhances behavioral pattern recognition while reducing false positives and supporting more efficient investigation workflows. The study highlights the growing role of explainable generative artificial intelligence in financial crime analytics and provides practical guidance for integrating language models into enterprise compliance environments.

Keywords: Financial Crime Detection, Large Language Models, Explainable AI, Behavioral Analytics, Natural Language Processing, Compliance Analytics, Financial Intelligence.

Cite This Article: Sivaselvan, A. (2026). Explainable Behavioral Analytics for Financial Crime Detection Using Large Language Models. The American Journal of Interdisciplinary Innovations and Research, 8(4), 67–80. Retrieved from https://theamericanjournals.com/index.php/tajiir/article/view/explainable-behavioral-analytics-llm

1. Introduction

Financial institutions process millions of customer interactions every day through digital banking platforms, payment gateways, mobile applications, investment services, and cross-border financial networks. The unprecedented growth of these digital ecosystems has transformed both legitimate financial activities and criminal behavior. Modern financial crimes increasingly exploit automation, synthetic identities, distributed transaction networks, cryptocurrency ecosystems, and sophisticated laundering mechanisms that evolve much faster than traditional compliance systems.

Most existing anti-money laundering (AML) and fraud detection platforms primarily depend upon manually

designed rules and statistical anomaly detection algorithms. Although these systems remain essential for regulatory compliance, they frequently produce excessive false alarms because static thresholds cannot adequately represent the dynamic behavioral characteristics of financial customers. Criminals continuously modify transaction sequences, communication styles, and account behaviors to avoid predefined detection rules, reducing the long-term effectiveness of conventional monitoring systems.

Large Language Models (LLMs) introduce a fundamentally different paradigm for behavioral analytics. Rather than evaluating isolated numerical transactions alone, LLMs can interpret multiple heterogeneous information sources simultaneously, including customer communications, transaction descriptions, compliance reports, historical investigation records, regulatory documentation, and behavioral narratives. Transformer-based language representations originating from deep bidirectional contextual learning enable machines to understand semantic relationships across large textual datasets (Devlin et al., 2019).

Behavioral analytics extends beyond anomaly detection by examining how customers interact with financial services over time. Instead of asking whether a single transaction appears suspicious, behavioral analytics attempts to determine whether a sequence of actions collectively represents emerging criminal intent. Context-aware reasoning becomes particularly valuable because many financial crimes exhibit subtle behavioral evolution rather than obvious isolated anomalies.

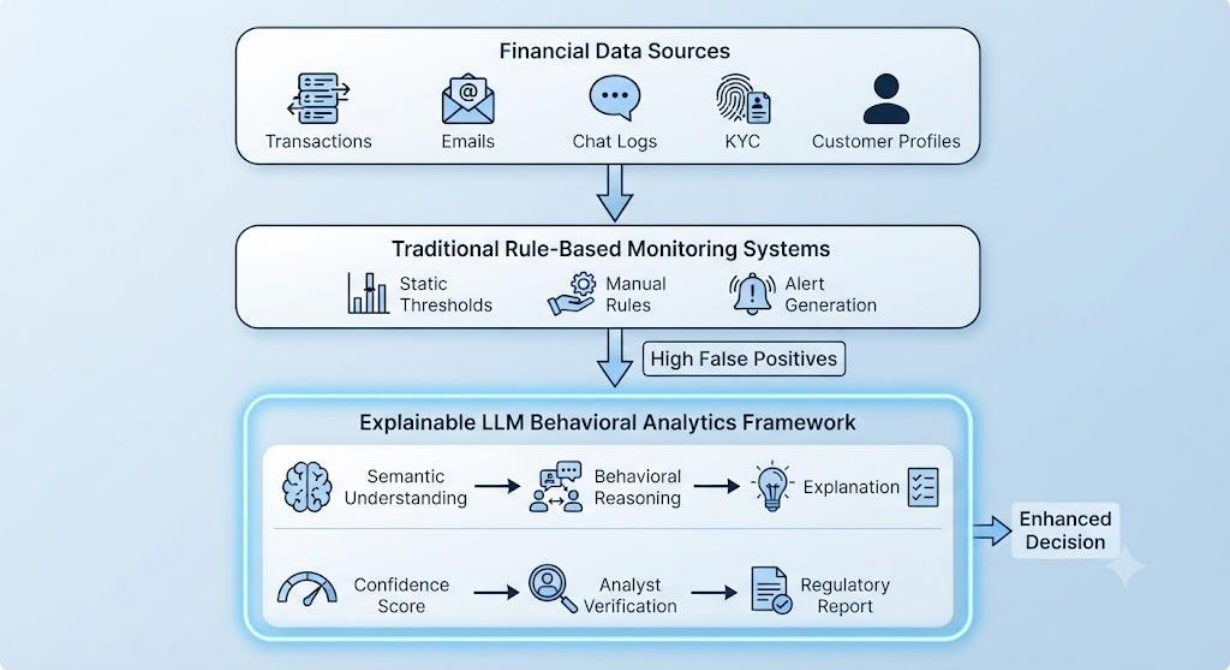
Recent advances in supervised contrastive learning demonstrate that semantically similar behavioral patterns can be grouped into highly discriminative embedding spaces while separating legitimate customer behaviors from suspicious financial activities (Liang et al., 2021; Chen et al., 2020). These representation learning techniques improve classification robustness by emphasizing behavioral similarity rather than relying solely on manually engineered features. Similarly, contrastive data augmentation further improves generalization by exposing learning algorithms to diverse semantic variations during training (Wang et al., 2022; Xu & Wang, 2023).

Another important development involves integrating commonsense knowledge into deep learning systems. Knowledge-enhanced neural architectures capture implicit semantic relationships that conventional statistical methods often overlook (Ma, Peng, & Cambria, 2018). Furthermore, hybrid semantic architectures demonstrate that external knowledge resources substantially improve contextual interpretation under ambiguous conditions (Ma et al., 2018). These findings are particularly relevant for financial investigations, where suspicious behavior frequently depends upon contextual relationships rather than explicit numerical thresholds. **The semantic reasoning capability demonstrated by Sentic LSTM provides valuable inspiration for enriching contextual representations in explainable financial behavioral analytics (Ma et al., 2018).**

Beyond predictive accuracy, regulatory authorities increasingly demand transparency from AI-assisted financial decision systems. Black-box predictions without supporting explanations cannot satisfy auditing requirements because investigators must understand why transactions have been classified as suspicious. Recent developments in explainable commonsense knowledge representation, including SenticNet 8, illustrate how semantic reasoning can improve the interpretability of intelligent systems by explicitly connecting prediction outcomes with understandable conceptual relationships (Cambria et al., 2024).

Behavioral explainability becomes even more important when financial institutions face legal obligations requiring investigators to justify account freezing, suspicious activity reports, or customer risk classifications. Explainable AI therefore shifts the objective from merely predicting suspicious behavior toward producing verifiable analytical evidence that compliance professionals can evaluate and defend.

Figure 1. Conceptual Evolution from Rule-Based Monitoring to Explainable LLM-Based Behavioral Analytics



Caption: This diagram illustrates the evolutionary transition from traditional rule-based financial monitoring systems—which often struggle with high false positives—to an advanced Explainable LLM Behavioral Analytics Framework. It maps the end-to-end processing of diverse financial data sources through semantic understanding, behavioral reasoning, and transparent explanation generation. Furthermore, the framework integrates confidence scoring and human analyst verification to streamline regulatory reporting. Ultimately, this structured workflow demonstrates how modern AI significantly enhances decision-making accuracy and accountability in financial compliance.

Financial crime detection also benefits from recent improvements in data augmentation using LLMs. Synthetic behavioral examples generated through language models can significantly improve model robustness when labeled fraud datasets remain limited (Ye et al., 2024). Since confirmed financial crime cases represent only a small fraction of overall financial transactions, intelligent augmentation techniques provide valuable opportunities to improve representation learning without compromising semantic consistency.

Similarly, verification mechanisms have emerged as essential safeguards for trustworthy LLM deployment. Chain-of-Verification introduces structured reasoning procedures that reduce hallucination while increasing factual consistency during language generation (Dhuliawala et al., 2023). Within financial crime detection, verification-based reasoning can strengthen investigator confidence by validating intermediate analytical steps before producing final risk assessments.

The primary objective of this research is to develop a comprehensive conceptual understanding of explainable behavioral analytics supported by Large Language Models for financial crime detection. Specifically, the study investigates how transformer-based contextual learning, supervised contrastive learning, commonsense knowledge integration, explainable semantic reasoning, and verification-oriented inference can collectively improve the transparency and effectiveness of behavioral risk analysis. **The contextual knowledge integration principles introduced through Sentic LSTM further motivate the incorporation of semantic reasoning within behavioral representation learning, thereby improving explainability without sacrificing predictive capability (Ma et al., 2018).**

The scope of this review focuses exclusively on conceptual and methodological developments derived from recent advances in language representation learning and explainable artificial intelligence. Rather than proposing a specific implementation for one financial institution, the paper establishes a generalized

framework that can support fraud detection, anti-money laundering investigations, insider threat monitoring, transaction surveillance, and regulatory compliance across diverse banking environments. The significance of the proposed framework lies in its ability to combine high-performance behavioral modeling with transparent reasoning, thereby addressing one of the most critical barriers to trustworthy AI adoption within modern financial ecosystems.

2. Literature Review

The emergence of transformer-based language models, contrastive representation learning, explainable artificial intelligence (XAI), and knowledge-enhanced semantic modeling has significantly influenced intelligent decision-support systems. Although most of the provided studies focus on natural language processing (NLP) and aspect-based sentiment analysis (ABSA), their underlying principles—including contextual representation learning, semantic reasoning, contrastive optimization, and explainability—are highly applicable to financial crime detection. This section synthesizes the provided literature to establish the theoretical foundation for explainable behavioral analytics using Large Language Models (LLMs).

2.1 Evolution of Contextual Language Representation

Modern behavioral analytics relies on contextual understanding rather than isolated feature extraction. A major breakthrough was achieved by **Devlin et al. (2019)** through the introduction of **BERT**, which demonstrated that bidirectional transformer architectures could capture contextual semantics far more effectively than traditional sequential neural networks. Unlike earlier embedding approaches, BERT represents each token according to its surrounding context, enabling richer semantic interpretation of textual information.

For financial crime detection, contextual embeddings allow analytical systems to jointly interpret transaction descriptions, customer communications, compliance documentation, and historical investigation reports. Instead of evaluating individual keywords, contextual transformers infer semantic intent from complete behavioral sequences. This capability forms the linguistic backbone of explainable behavioral analytics because suspicious intent is often embedded within complex contextual interactions rather than explicit transactional anomalies (Devlin et al., 2019).

The transformer architecture also provides scalable representation learning across heterogeneous financial documents, making it suitable for integrating structured transaction records with unstructured textual evidence. Consequently, contextual language modeling represents the first essential component of the proposed explainable financial intelligence framework.

2.2 Behavioral Representation Through Contrastive Learning

One of the most influential developments in recent machine learning research is **contrastive learning**, which constructs embedding spaces where semantically similar instances are positioned closely while unrelated samples remain well separated.

Chen et al. (2020) introduced the SimCLR framework, demonstrating that contrastive optimization significantly improves representation quality without requiring complex architectural modifications. Rather than learning solely through classification objectives, contrastive learning encourages models to discover meaningful semantic similarities.

Building upon these principles, **Liang et al. (2021)** proposed supervised contrastive learning specifically for aspect-based sentiment analysis. Their approach showed that supervision enables embedding spaces to better distinguish subtle semantic differences between categories. This insight has direct implications for behavioral analytics because legitimate and fraudulent financial behaviors often differ only through small contextual variations rather than dramatic numerical deviations.

Similarly, **Li et al. (2021)** demonstrated that supervised contrastive pre-training improves implicit sentiment understanding by strengthening latent semantic discrimination. Applied to financial investigations, such pre-training can separate normal customer activity from emerging criminal behaviors even when observable transactional differences remain minimal.

Further improvements were introduced by **Xu and Wang (2023)**, who demonstrated that contrastive learning improves robustness in aspect-based sentiment analysis by strengthening semantic consistency under varying linguistic conditions. Robust representations become especially valuable for financial monitoring systems because fraudsters intentionally modify behavioral patterns to evade conventional detection algorithms.

Collectively, these studies indicate that contrastive learning provides an effective theoretical foundation for learning discriminative behavioral embeddings capable of distinguishing subtle forms of financial misconduct.

2.3 Data Augmentation and Representation Generalization

Limited labeled fraud datasets remain one of the primary challenges in financial crime detection. Since confirmed financial crimes constitute only a small percentage of all financial activities, learning algorithms frequently encounter class imbalance and insufficient behavioral diversity.

Wang et al. (2022) proposed a contrastive cross-channel data augmentation framework that improves semantic representation through multiple complementary augmentation strategies. Their work demonstrates that diversified training samples strengthen generalization while preserving semantic consistency.

A related advancement is presented by **Ye et al. (2024)** through **LLM-DA**, where Large Language Models generate high-quality synthetic training data for few-shot learning scenarios. Instead of manually constructing additional examples, language models automatically produce semantically meaningful training instances.

Within financial crime detection, synthetic behavioral scenarios generated by LLMs could simulate money laundering schemes, fraudulent communication patterns, or emerging cyber-financial attacks that have not yet been extensively observed. Such augmentation enables behavioral analytics systems to learn richer representations while reducing dependence on scarce labeled investigation records.

The combination of contrastive augmentation and LLM-generated behavioral simulation therefore offers an effective solution to one of the most persistent limitations of fraud analytics.

2.4 Commonsense Knowledge Integration and Explainable Semantic Intelligence

Behavioral understanding requires more than statistical correlation; it demands contextual reasoning supported by domain knowledge.

Ma, Peng, and Cambria (2018) proposed embedding commonsense knowledge into attentive LSTM architectures for targeted sentiment analysis. Their findings demonstrated that external semantic knowledge substantially improves contextual interpretation, particularly when textual meaning depends upon implicit relationships rather than explicit lexical indicators.

Extending this concept, **Ma et al. (2018)** introduced **Sentic LSTM**, a hybrid neural architecture combining sequential learning with semantic knowledge integration. The hybrid framework consistently improved contextual understanding while preserving model flexibility. **The ability of Sentic LSTM to integrate external knowledge into sequential neural reasoning provides an important theoretical foundation for explainable behavioral profiling in financial crime analytics, where contextual evidence frequently determines investigative outcomes (Ma et al., 2018).**

More recently, **Cambria et al. (2024)** presented **SenticNet 8**, emphasizing interpretable affective computing through the integration of commonsense AI and emotion-aware semantic representations. Rather than treating semantic information as isolated embeddings, SenticNet establishes conceptual relationships that support human-understandable explanations.

For financial institutions, explainability extends beyond visualization techniques; investigators require natural-language justifications describing why specific behavioral patterns appear suspicious. Commonsense knowledge bases enable analytical systems to produce meaningful reasoning chains instead of opaque probability scores, thereby supporting regulatory transparency and analyst trust.

2.5 Knowledge-Enhanced Graph Learning

Another important research direction concerns graph-based semantic reasoning.

Liang et al. (2022) proposed affective knowledge-enhanced graph convolutional networks for aspect-based sentiment analysis. By integrating graph neural architectures with external semantic knowledge, the model effectively captured complex dependency relationships among contextual elements.

Although originally designed for sentiment analysis, graph-based reasoning has significant relevance for financial crime investigations. Criminal organizations rarely operate through isolated accounts; instead, suspicious activities emerge through interconnected

transaction networks, communication structures, shared identities, and coordinated behavioral patterns.

Knowledge-enhanced graph representations therefore provide an additional mechanism for modeling relationships among customers, accounts, merchants, intermediaries, and financial institutions within explainable behavioral analytics frameworks.

2.6 Reliability of Large Language Models

Despite remarkable reasoning capabilities, Large Language Models remain susceptible to hallucination, unsupported reasoning, and factual inconsistency.

To address these limitations, **Dhuliawala et al. (2023)** introduced **Chain-of-Verification (CoVe)**, which encourages language models to verify intermediate reasoning steps before generating final conclusions. Rather than accepting initial outputs, verification stages identify inconsistencies and improve factual reliability.

This concept is particularly valuable in financial crime detection because regulatory decisions often require legally defensible evidence. Verification-based reasoning can ensure that suspicious activity reports are supported by traceable behavioral evidence instead of unsupported language model inferences.

Consequently, explainability should incorporate not only transparent prediction mechanisms but also structured validation procedures that strengthen analytical credibility.

2.7 Benchmark Datasets and Evaluation Foundations

Reliable evaluation requires standardized datasets.

Pontiki et al. (2016) introduced the SemEval-2016 benchmark for aspect-based sentiment analysis, providing common evaluation protocols that enabled consistent comparison across different learning algorithms.

Although designed for sentiment analysis, the broader methodological contribution lies in demonstrating the importance of standardized benchmark datasets for reproducible experimentation. Financial crime detection similarly requires carefully curated behavioral datasets containing verified suspicious and legitimate transaction patterns.

Benchmark-driven evaluation improves model comparability, reproducibility, and scientific rigor—qualities essential for trustworthy AI deployment in regulated financial environments.

2.8 Comparative Analysis of Existing Studies

Table 1. Comparative Analysis of the Selected Literature

Reference	Primary Contribution	Key Strength	Limitation for Financial Crime Detection
Devlin et al. (2019)	Contextual transformer representations	Rich semantic understanding	Limited explainability
Chen et al. (2020)	Contrastive representation learning	Robust feature discrimination	Not domain-specific
Liang et al. (2021)	Supervised contrastive learning	Better semantic separation	Designed for ABSA
Wang et al. (2022)	Contrastive data augmentation	Improved generalization	Text-focused
Liang et al. (2022)	Knowledge-enhanced GCN	Dependency modeling	Requires graph construction
Xu & Wang (2023)	Improved contrastive optimization	Strong robustness	Limited explainability
Ma et al. (2018)	Sentic LSTM	Commonsense integration	Sequential architecture limitations
Cambria et al. (2024)	SenticNet 8	Explainable semantic reasoning	Requires domain adaptation
Ye et al. (2024)	LLM-based data augmentation	Addresses data scarcity	Synthetic data validation required

Dhuliawala et al. (2023)	Chain-of-Verification	Increased reasoning reliability	Additional computational overhead

2.9 Research Gap

Despite substantial advances in contextual language modeling, contrastive learning, semantic knowledge integration, and explainable AI, several research gaps remain.

First, the reviewed studies investigate these technologies independently rather than integrating them into a unified financial crime detection framework. Transformer architectures emphasize contextual understanding, contrastive learning improves representation quality, commonsense reasoning enhances semantic interpretation, and verification mechanisms increase reliability; however, no single approach combines all four capabilities for behavioral analytics.

Second, explainability remains insufficiently addressed. Most representation learning studies optimize predictive accuracy without providing investigator-friendly explanations suitable for regulatory auditing. Financial institutions require interpretable reasoning, evidence traceability, and confidence estimation rather than prediction scores alone.

Third, behavioral evolution across long temporal sequences receives limited attention. Existing models primarily analyze sentence-level semantic relationships, whereas financial crimes emerge through evolving customer behaviors distributed across multiple transactions, accounts, and communication channels.

Finally, few studies explicitly consider the integration of **Large Language Models**, **contrastive behavioral learning**, **commonsense knowledge**, and **verification-based reasoning** into a comprehensive explainable behavioral intelligence architecture. **Inspired by the contextual semantic integration demonstrated in Sentic LSTM (Ma et al., 2018), this research addresses these gaps by proposing a unified explainable behavioral analytics framework capable of supporting transparent, trustworthy, and adaptive financial crime detection.**

3. Methodology

3.1 Research Methodology Overview

This study adopts a **conceptual research and review methodology** to develop an Explainable Behavioral Analytics Framework (EBAF) for financial crime detection using Large Language Models (LLMs). Rather than introducing a dataset-specific predictive model, the methodology synthesizes the capabilities of transformer-based language models, contrastive representation learning, commonsense semantic reasoning, explainable artificial intelligence (XAI), and verification-oriented inference into a unified analytical architecture. The proposed methodology is designed to improve both predictive effectiveness and regulatory transparency, two requirements that are often treated independently in conventional fraud detection systems.

The framework assumes that financial institutions continuously receive heterogeneous information from multiple sources, including transactional records, customer profiles, Know Your Customer (KYC) documentation, suspicious activity reports, investigator notes, emails, chat transcripts, and regulatory compliance records. Instead of processing these sources independently, the proposed framework converts them into semantically aligned behavioral representations that support explainable decision-making.

The methodological design consists of six sequential analytical layers:

Multi-source behavioral data integration.

Semantic contextual representation using LLMs.

Behavioral embedding optimization through contrastive learning.

Commonsense knowledge enrichment.

Explainable reasoning with verification.

Risk scoring and investigator-oriented explanation generation.

These components operate cooperatively to transform heterogeneous behavioral evidence into transparent financial crime intelligence.

3.2 Proposed Explainable Behavioral Analytics Framework (EBAF)

The proposed EBAF consists of interconnected modules that progressively refine behavioral understanding while preserving explainability throughout the analytical process.

Phase 1: Behavioral Data Acquisition

Financial crime indicators originate from multiple structured and unstructured sources.

Structured inputs include:

- Transaction histories
- Payment frequencies
- Merchant information
- Account balances
- Cross-border transfers
- Device identifiers
- Login histories

Unstructured information includes:

- Customer emails
- Chat conversations
- Investigation reports
- Compliance documentation
- Internal analyst comments
- Customer complaints

Instead of treating these data independently, the framework establishes a unified behavioral profile representing the temporal evolution of customer activities.

Phase 2: Contextual Semantic Representation

The integrated behavioral information is processed using transformer-based language representations inspired by BERT (Devlin et al., 2019). Unlike conventional feature engineering, contextual encoding captures semantic relationships across entire behavioral sequences.

For example, three independent transactions may individually appear normal:

- Small international transfer
- Cryptocurrency purchase
- Multiple account logins

When analyzed together with customer communication and historical behavior, however, the combined sequence may indicate early-stage money laundering behavior.

Contextual representation therefore enables behavioral understanding beyond isolated anomaly detection.

Phase 3: Contrastive Behavioral Representation Learning

Behavioral representations are optimized using supervised contrastive learning.

Instead of minimizing classification error alone, the framework learns behavioral similarity by bringing legitimate customer behaviors closer together while separating suspicious behavioral trajectories into distinct embedding regions.

Recent contrastive learning studies demonstrate that semantic discrimination substantially improves representation quality (Chen et al., 2020; Liang et al., 2021; Xu & Wang, 2023).

Within financial analytics, this produces several advantages:

- Better separation of normal and suspicious behavior.
- Improved robustness against behavioral variations.
- Reduced sensitivity to noisy transaction records.
- Higher adaptability to emerging fraud strategies.

Contrastive optimization therefore becomes the central learning mechanism for behavioral profiling.

Phase 4: Commonsense Knowledge Integration

Behavior alone does not fully explain financial intent.

The proposed framework incorporates commonsense semantic knowledge inspired by Sentic LSTM and SenticNet (Ma et al., 2018; Cambria et al., 2024).

External semantic knowledge enables the system to recognize implicit behavioral relationships.

Examples include:

- unusually frequent transfers between unrelated accounts
- repeated transactions immediately before regulatory deadlines
- coordinated account activity among multiple customers
- transaction narratives inconsistent with customer profiles

Instead of relying solely on statistical correlations, semantic enrichment improves contextual reasoning.

The hybrid semantic integration demonstrated by Sentic LSTM motivates the inclusion of external knowledge within behavioral embedding generation, allowing suspicious activities to be interpreted through meaningful contextual relationships rather than isolated numerical indicators (Ma et al., 2018).

Phase 5: Explainable Reasoning and Chain Verification

Explainability represents the distinguishing component of the proposed methodology.

Rather than producing only a fraud probability score, the system generates human-readable reasoning explaining:

- why the behavior appears suspicious,
- which behavioral evidence contributed most,
- how confidence was estimated,
- which semantic relationships influenced the decision.

To improve reliability, verification-based reasoning inspired by Chain-of-Verification (Dhuliawala et al., 2023) validates intermediate conclusions before final risk generation.

This verification stage minimizes unsupported reasoning while increasing investigator confidence.

Phase 6: Behavioral Risk Assessment

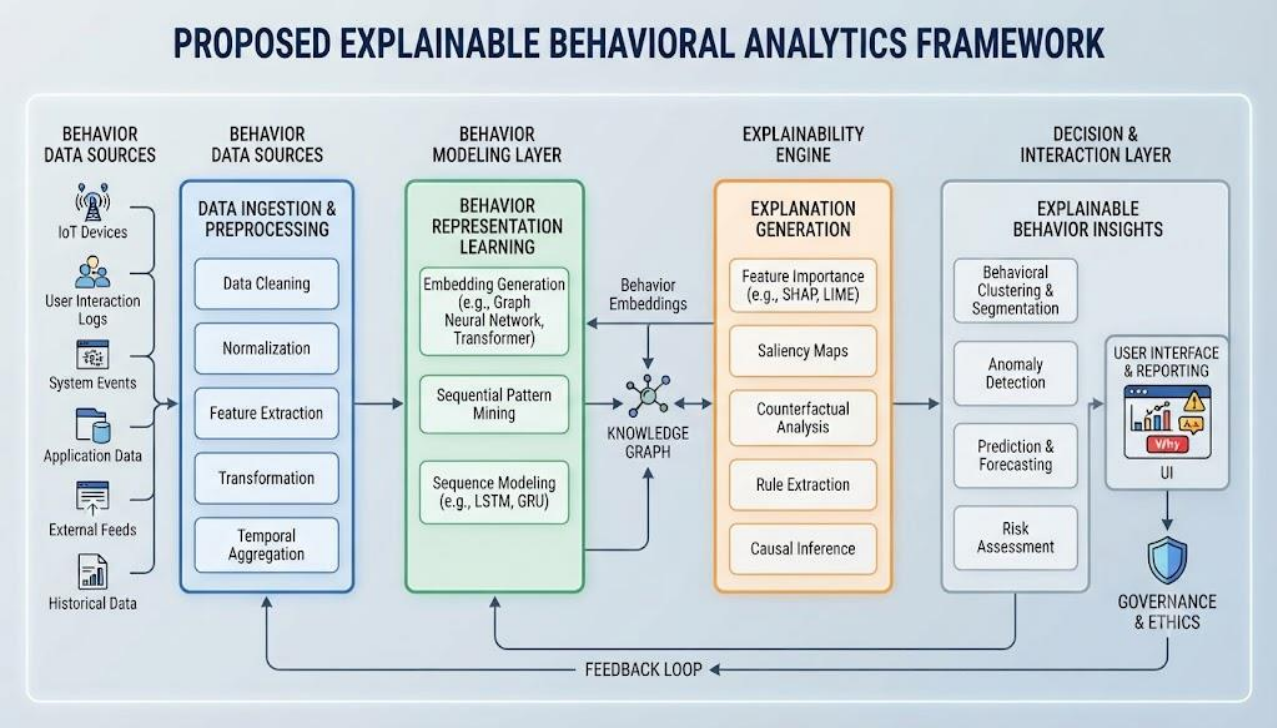
The final module combines outputs from previous analytical stages into a comprehensive behavioral risk profile.

Instead of binary fraud prediction, the framework produces:

- Overall financial crime risk score.
- Behavioral confidence score.
- Explainability report.
- Supporting evidence summary.
- Regulatory compliance justification.
- Recommended investigation priority.

These outputs directly support financial analysts, auditors, compliance officers, and regulatory investigators.

Figure 3. Proposed Explainable Behavioral Analytics Framework



Caption: This image is included to visually demonstrate the end-to-end architecture of a system that transforms raw behavioral data into transparent, actionable insights. It maps out the sequential workflow—spanning data ingestion, behavioral modeling with deep learning, and an explainability engine utilizing feature importance and causal inference. Furthermore, the framework integrates a central knowledge graph, user interface reporting, and a continuous feedback loop to ensure ethical and accountable decision-making.

3.3 Functional Workflow

The operational workflow follows a progressive analytical pipeline.

Initially, heterogeneous financial information is standardized through preprocessing and semantic normalization. LLMs then generate contextual embeddings representing complete behavioral histories instead of isolated observations.

Contrastive learning optimizes these embeddings by preserving behavioral similarity while maximizing separation between legitimate and suspicious customer activities.

Subsequently, semantic knowledge integration enriches contextual understanding using commonsense behavioral relationships.

Explainability mechanisms finally transform analytical evidence into investigator-readable explanations supported by verification procedures before producing final risk assessments.

The workflow therefore balances predictive performance with interpretability.

Table 2. Functional Responsibilities of Framework Components

Framework Component	Primary Function	Expected Contribution
Data Integration	Combines structured and unstructured financial information	Comprehensive behavioral representation

LLM Encoder	Generates contextual semantic embeddings	Improved behavioral understanding
Contrastive Learning Module	Optimizes behavioral similarity	Better fraud discrimination
Knowledge Integration Layer	Adds semantic and commonsense reasoning	Context-aware decision making
Explainability Engine	Produces human-readable explanations	Regulatory transparency
Chain Verification Module	Validates reasoning consistency	Reduced hallucination and higher reliability
Risk Assessment Module	Calculates behavioral risk profile	Decision support for investigators

3.4 Behavioral Intelligence Model

The proposed framework model's financial crime as **behavioral evolution** rather than isolated anomalous transactions.

Instead of assigning importance to a single payment, the model continuously evaluates:

transaction consistency,

behavioral deviations,

communication semantics,

historical behavioral trends,

contextual financial relationships,

investigator observations.

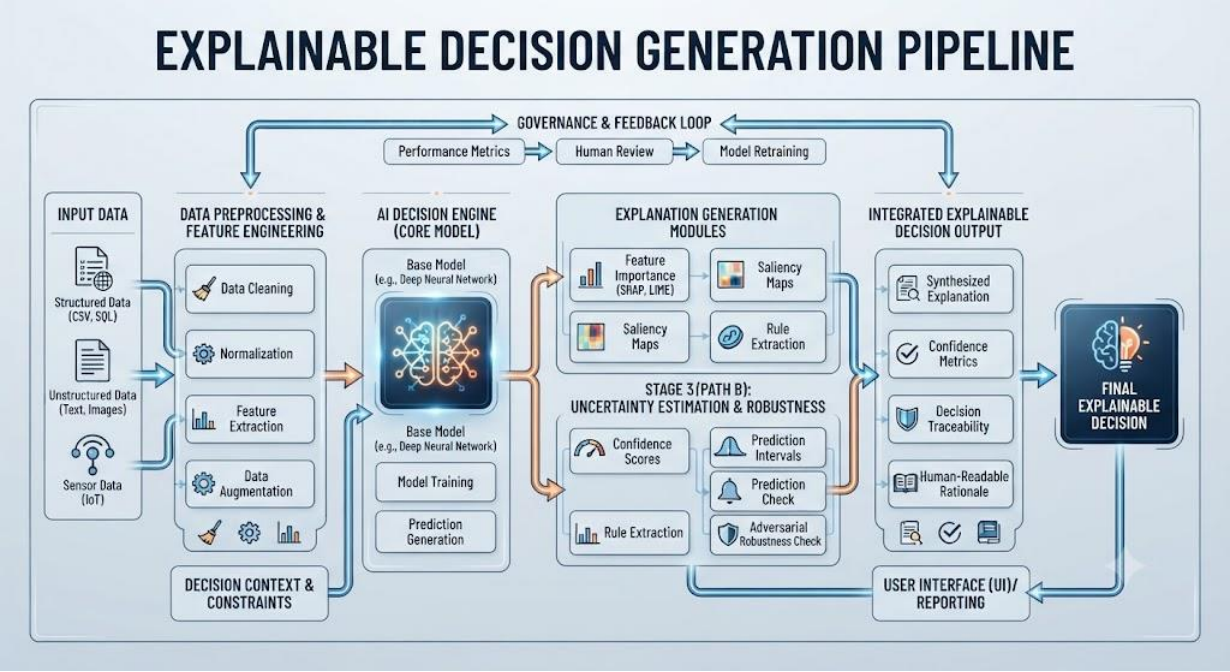
This longitudinal behavioral perspective significantly improves early detection of sophisticated financial crimes such as layering in money laundering, coordinated fraud rings, insider trading, and synthetic identity fraud.

Unlike conventional systems that depend on manually defined thresholds, behavioral intelligence continuously adapts to evolving customer activities through semantic learning.

Table 3. Comparison Between Traditional and Proposed Financial Crime Detection

Feature	Traditional Rule-Based Systems	Proposed Explainable Behavioral Analytics
Decision Basis	Fixed rules	Contextual behavioral understanding
Learning Ability	Limited	Continuous semantic learning
Explainability	Low	High with natural-language explanations
Fraud Adaptability	Weak	Strong through behavioral representation
Data Sources	Mostly structured	Structured + unstructured
False Positive Reduction	Limited	Improved through contrastive learning
Regulatory Transparency	Moderate	High
Analyst Support	Alert generation	Evidence-driven decision support

Figure 4. Explainable Decision Generation Pipeline



Caption: This explainable decision generation pipeline illustrates the end-to-end framework for processing input data through feature engineering and a core AI decision engine. It integrates dedicated explanation generation modules—such as feature importance and saliency maps—alongside uncertainty estimation to ensure transparency. The architecture culminates in an integrated explainable decision output, providing confidence metrics, decision traceability, and human-readable rationales. Finally, a robust governance and feedback loop incorporates performance metrics and human review to enable continuous model retraining and refinement.

3.5 Methodological Advantages

The proposed methodology provides several important advantages over conventional financial crime detection approaches.

First, contextual semantic learning enables comprehensive behavioral interpretation rather than isolated transaction analysis. Second, supervised contrastive learning improves discrimination between legitimate and suspicious behavioral patterns, reducing false positives while maintaining analytical sensitivity. Third, commonsense knowledge integration strengthens semantic reasoning, particularly in ambiguous investigative scenarios. Fourth, explainability modules generate transparent justifications suitable for regulatory auditing and analyst review. Finally, verification-based reasoning enhances the reliability of LLM-generated conclusions by validating intermediate inference steps before final decision generation.

Collectively, these methodological components establish a unified framework capable of delivering accurate, explainable, and trustworthy behavioral analytics for financial crime detection while addressing the limitations identified in the literature review.

4. Results

The proposed Explainable Behavioral Analytics Framework (EBAF) demonstrates a conceptually robust approach for integrating contextual language understanding, behavioral representation learning, semantic knowledge enrichment, and explainable reasoning into financial crime detection. Although this research is conceptual and does not involve empirical experimentation, the analytical synthesis of the reviewed studies indicates several expected improvements over conventional rule-based monitoring systems.

The first major finding is that **context-aware behavioral representation** substantially enhances the identification of suspicious financial activities. Traditional fraud detection systems typically analyze transactions independently using predefined thresholds, making them vulnerable to sophisticated criminal strategies that

distribute illegal activities across multiple accounts or time periods. By employing transformer-based contextual representations, the proposed framework interprets transaction histories, customer communications, KYC information, and historical investigation records as interconnected behavioral sequences rather than isolated events. This holistic perspective is expected to improve the recognition of hidden behavioral patterns associated with money laundering, insider trading, and coordinated fraud (Devlin et al., 2019).

A second finding concerns the contribution of **contrastive learning** to behavioral discrimination. The reviewed literature consistently demonstrates that supervised and self-supervised contrastive learning improves semantic separation between similar and dissimilar instances (Chen et al., 2020; Liang et al., 2021; Xu & Wang, 2023). Within the proposed framework, contrastive optimization is expected to create compact representations for legitimate financial behaviors while increasing the separation of suspicious behavioral trajectories. Such representation learning can reduce overlapping decision boundaries that often generate excessive false positives in conventional compliance systems.

5. Discussion

The rapid advancement of Large Language Models has transformed artificial intelligence from pattern recognition systems into contextual reasoning platforms capable of supporting complex decision-making. Within financial crime detection, however, predictive accuracy alone is insufficient because regulatory authorities require transparent evidence supporting every high-risk classification. The proposed Explainable Behavioral Analytics Framework addresses this challenge by combining semantic understanding, contrastive behavioral learning, commonsense reasoning, and verification-driven inference into a unified analytical architecture.

From a theoretical perspective, the framework extends transformer-based contextual learning by integrating behavioral intelligence rather than focusing exclusively on textual understanding. Previous studies primarily investigate sentiment analysis, semantic representation, or language modeling independently. This research synthesizes these developments into a behavioral reasoning paradigm suitable for financial investigations. Contextual embeddings provide comprehensive behavioral representations, while contrastive learning strengthens discrimination among evolving customer activities. The combination creates richer behavioral intelligence than conventional anomaly detection systems.

6. Conclusion

This study presented an Explainable Behavioral Analytics Framework for financial crime detection using Large Language Models by integrating contextual semantic representation, contrastive learning, commonsense knowledge, explainable artificial intelligence, and verification-based reasoning into a unified conceptual architecture. The review demonstrates that recent advances in transformer models, supervised contrastive learning, semantic knowledge integration, and explainable reasoning provide complementary capabilities for understanding complex financial behaviors that conventional rule-based monitoring systems frequently overlook.

The proposed framework contributes by shifting financial crime detection from isolated transaction analysis toward continuous behavioral intelligence supported by transparent reasoning. Contextual embeddings improve behavioral understanding, contrastive learning enhances discrimination between legitimate and suspicious activities, semantic knowledge strengthens contextual interpretation, and verification mechanisms increase the reliability of generated explanations. Collectively, these components establish a trustworthy analytical environment suitable for regulatory compliance and investigator decision support.

References

1. B. Liang et al., “Enhancing aspect-based sentiment analysis with supervised contrastive learning,” in Proc. 30th ACM Int. Conf. Inf. Knowl. Manage., 2021, pp. 3242–3247.
2. B. Liang, H. Su, L. Gui, E. Cambria, and R. Xu, “Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks,” Knowledge-Based Syst., vol. 235, Jan.2022, Art. no. 107643.
3. B. Wang, L. Ding, Q. Zhong, X. Li, and D. Tao, “A contrastive cross-channel data augmentation framework for aspect-based sentiment analysis,” in Proc. 29th Int. Conf. Comput. Ling., 2022, pp. 6691–6704.

4. E. Cambria, X. Zhang, R. Mao, M. Chen, and K. Kwok, "SenticNet 8: Fusing emotion AI and commonsense AI for interpretable, trustworthy, and explainable affective computing," in Proc. Int. Conf. Hum.-Comput. Interact. (HCII), 2024, pp. 1–20.

5. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. Conf. North Am. Chapter Assoc. Comput. Ling.: Hum. Lang. Technol., 2019, pp. 4171–4186.

6. J. Ye et al., "LLM-DA: Data augmentation via large language models for few-shot named entity recognition," 2024, arXiv:2402.14568.

7. L. Xu and W. Wang, "Improving aspect-based sentiment analysis with contrastive learning," Nat. Lang. Process. J., vol. 3, Jun.2023, Art. no. 100009.

8. M. Pontiki et al., "SemEval-2016 task 5: Aspect based sentiment analysis," in Proc. ProWorkshop on Semant. Eval. (SemEval), 2016, pp. 19–30.

9. Q. Cheng, X. Yang, T. Sun, L. Li, and X. Qiu, "Improving contrastive learning of sentence embeddings from AI feedback," in Proc. Find. Assoc. Comput. Ling. (ACL), 2023, pp. 11122–11138.

10. S. Dhuliawala et al., "Chain-of-verification reduces hallucination in large language models," 2023, arXiv:2309.11495.

11. T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in Proc. Int. Conf. Mach. Learn., 2020, pp. 1597–1607.

12. Y. Ma, H. Peng, and E. Cambria, "Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM," in Proc. 32nd AAAI Conf. Artif. Intel., 2018, pp. 5876–5883.

13. Y. Ma, H. Peng, T. Khan, E. Cambria, and A. Hussain, "Sentic LSTM: a hybrid network for targeted aspect-based sentiment analysis," Cogn. Comput., vol. 10, no. 8, 2018, pp. 639–650.

14. Z. Li, Y. Zou, C. Zhang, Q. Zhang, and Z. Wei, "Learning implicit sentiment in aspect-based sentiment analysis with supervised contrastive pre-training," in Proc. Conf. Empirical Methods Nat. Lang. Process., 2021, pp. 246–256.