Scholars Journal of Engineering and Technology
OPEN ACCESS
Abbreviated Key Title: Sch J Eng Tech
ISSN 2347-9523 (Print) | ISSN 2321-435X (Online)
Journal homepage: https://saspublishers.com

Neuro-Symbolic Artificial Intelligence for Secure, Explainable, and Intelligent Banking and FinTech Systems: Fraud Detection, Credit Risk, Financial Forecasting, and Transaction Security

Abu Bakar Razzaq[1], MD Zia Ul Haque[2], Sehrish Naz[3], Huzaim Razzaq[3], Saba Munawar[4]*, Noshaba Qasmi[5], Uzma Raza[6], Muhammad Noman[7], M. Zeeshan Younus[6]

[1]Department of Business Administration, Shah Abdul Latif University; Association of Chartered Certified Accountants (ACCA), Karachi, Sindh, Pakistan
[2]Department of Computer Science and Engineering, Northern University Bangladesh, Bangladesh
[3]Department of Business Administration, Iqra University, Karachi, Sindh, Pakistan
[4]Department of Computer Science, University of Agriculture Faisalabad, Faisalabad, Pakistan
[5]National Centre for Bioinformatics, Quaid-i-Azam University, Islamabad, Pakistan
[6]Faculty of Computer Science, Mohammad Ali Jinnah University, Karachi, Pakistan
[7]Department of Management Sciences, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST), Karachi, Sindh, Pakistan

DOI: https://doi.org/10.36347/sjet.2026.v14i11.001 | **Received:** 15.08.2026 | **Accepted:** 24.10.2026 | **Published:** 07.11.2026

***Corresponding author:** Saba Munawar
Department of Computer Science, University of Agriculture Faisalabad, Faisalabad, Pakistan

Abstract | **Original Research Article**

The rapid digitalization of banking and FinTech systems has intensified the need for artificial intelligence that is accurate, secure, explainable, and regulation-compliant. Conventional deep learning models deliver strong predictive performance but remain opaque black boxes, while Large Language Models (LLMs) offer powerful unstructured-data understanding yet suffer from hallucination and weak constraint satisfaction. Neither paradigm alone satisfies the transparency and compliance demands of modern financial regulation. This paper proposes the Explainable Neuro-Symbolic Trust Layer (XNSTL), a unified framework integrating LLM-based neural learning with symbolic rule reasoning, financial ontologies, and a regulatory audit mechanism to serve four critical banking tasks: fraud detection, credit risk assessment, financial forecasting, and transaction security. The XNSTL architecture comprises four modules: (i) an LLM-based feature extraction module for unstructured financial data; (ii) a symbolic rule engine encoding AML/KYC, Basel III, and institutional credit policies; (iii) a neuro-symbolic fusion module with confidence calibration and uncertainty quantification; and (iv) an explainability and audit layer generating counterfactual explanations and tamper-proof audit trails. We formalize the four tasks, define a hybrid loss function jointly optimizing predictive accuracy, symbolic constraint satisfaction, and explanation faithfulness, and propose an evaluation protocol using public benchmarks including IEEE-CIS Fraud Detection and Lending Club datasets. The framework is designed to reduce false positives in fraud detection, improve fairness in credit decisions, strengthen forecasting robustness under regime shifts, and enable real-time, audit-ready transaction security, establishing a pathway toward trustworthy and regulation-native financial AI.

Keywords: Neuro-Symbolic AI, Large Language Models, Fraud Detection, Credit Risk Assessment, Explainable AI, Fintech Security.

Copyright © 2026 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution **4.0 International License (CC BY-NC 4.0)** which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

1. INTRODUCTION

Banking and financial technology have become central to the modern economy. Digital payments, mobile banking, open banking, and block chain settlement systems now handle very large volumes of money every day [1, 2]. These systems depend on artificial intelligence for speed and accuracy. This section explains the background of these systems, the limits of current AI models, the need for neuro-symbolic and explainable AI, and the objectives of this study. Each part is presented in a separate subsection with a short introduction before it [3].

1.1 Background of Banking and FinTech Systems

Banking and financial technology have changed rapidly over the last decade. Digital payments, mobile banking, open banking platforms, and blockchain based settlement systems now handle very large volumes of money every day. A single large bank may process millions of transactions in a day, and global digital payment volumes continue to grow each year. Customers expect fast service, low cost, and strong security. Banks expect accurate risk control, low operational cost, and full compliance with regulations.

The growth of FinTech has also changed the structure of the financial market [4-10]. New entrants offer payments, lending, insurance, and investment services without the burden of traditional branch networks. These firms rely heavily on data and automation. They use mobile data, transaction history, and behavioural signals to make quick decisions. Traditional banks have responded by building their own digital channels and by partnering with FinTech firms. The result is a financial ecosystem that is faster, more connected, and more dependent on software than ever before [11].

This growth has also increased risk. Fraud, money laundering, identity theft, account takeover, and cyber attacks have become more common and more automated. Attackers now use scripting tools and bot networks to test thousands of accounts in a short time. Manual review teams cannot keep pace with this speed. Financial institutions have therefore turned to artificial intelligence for fraud detection, credit scoring, market forecasting, and transaction security.

The dependence on AI is now very deep. Every card swipe, loan application, fund transfer, and trading order passes through some model [12-17]. The quality of these models affects customer trust, institutional stability, and regulatory standing. A wrong fraud flag can block a genuine customer. A missed fraud can cause direct loss and reputational damage. A biased credit model can lead to legal action. A weak forecasting model can cause large trading losses. Because of this, the design of financial AI systems is no longer only a technical matter. It is also a matter of governance and public trust. Regulators have also increased their attention on AI in finance. Rules on data protection, fair lending, anti-money laundering, and model risk management now apply directly to automated decisions. Banks must be able to show how a decision was made, which data was used, and whether the decision followed policy. This creates a strong demand for AI systems that are not only accurate but also transparent and controllable [18].

1.2 Limitations of Current AI Models in Finance

Current AI models in banking show strong accuracy but weak transparency. Deep learning models work like black boxes. They learn complex patterns from large data sets and often outperform simpler models on prediction tasks. However, they cannot easily explain why a particular decision was made. A fraud alert or loan rejection often cannot be described in clear terms. Regulators and customers now demand reasons for such decisions. In many countries, customers have a legal right to know the main factors behind an automated decision. A second problem is compliance. Rule based systems follow regulations strictly but cannot learn from new data. Machine learning models learn from data but may break legal or policy rules. For example, a model may learn patterns that indirectly use protected attributes such as age or region. This can lead to unfair outcomes even when the protected attribute is not given as input. Combining strict rules with flexible learning is difficult with current designs. Most systems keep the two parts separate, and the rule layer is applied only after the model has already produced a decision. A third problem is security. Large models can be attacked with poisoned training data or with carefully crafted inputs at prediction time. In finance, such attacks can cause direct financial loss. Attackers may also try to reverse engineer a model to learn customer data or to find weak points. Large language models add another risk. They can produce confident but incorrect statements, a problem known as hallucination. Using such models without control in a regulated environment is dangerous. A fourth problem is maintenance. Financial rules change often. Tax rules, reporting thresholds, and credit policies are updated regularly. In a pure machine learning system, each change requires retraining and full validation. In a pure rule system, each change requires manual code updates. Neither approach handles change smoothly. A better design would allow rules and learned patterns to work together and to be updated without breaking the other part [19, 20].

Table 1: Limitations of Current AI Models in Banking and FinTech

Model Type	Main Strength	Key Limitation
Rule-based systems	Strict compliance	No learning from new data
Classical machine learning	Good accuracy on tabular data	Limited handling of complex patterns
Deep learning models	High accuracy on large data	Black box, hard to explain
Large language models	Strong text understanding	Hallucination and weak rule control
Hybrid early models	Partial transparency	Rules and learning stay separate

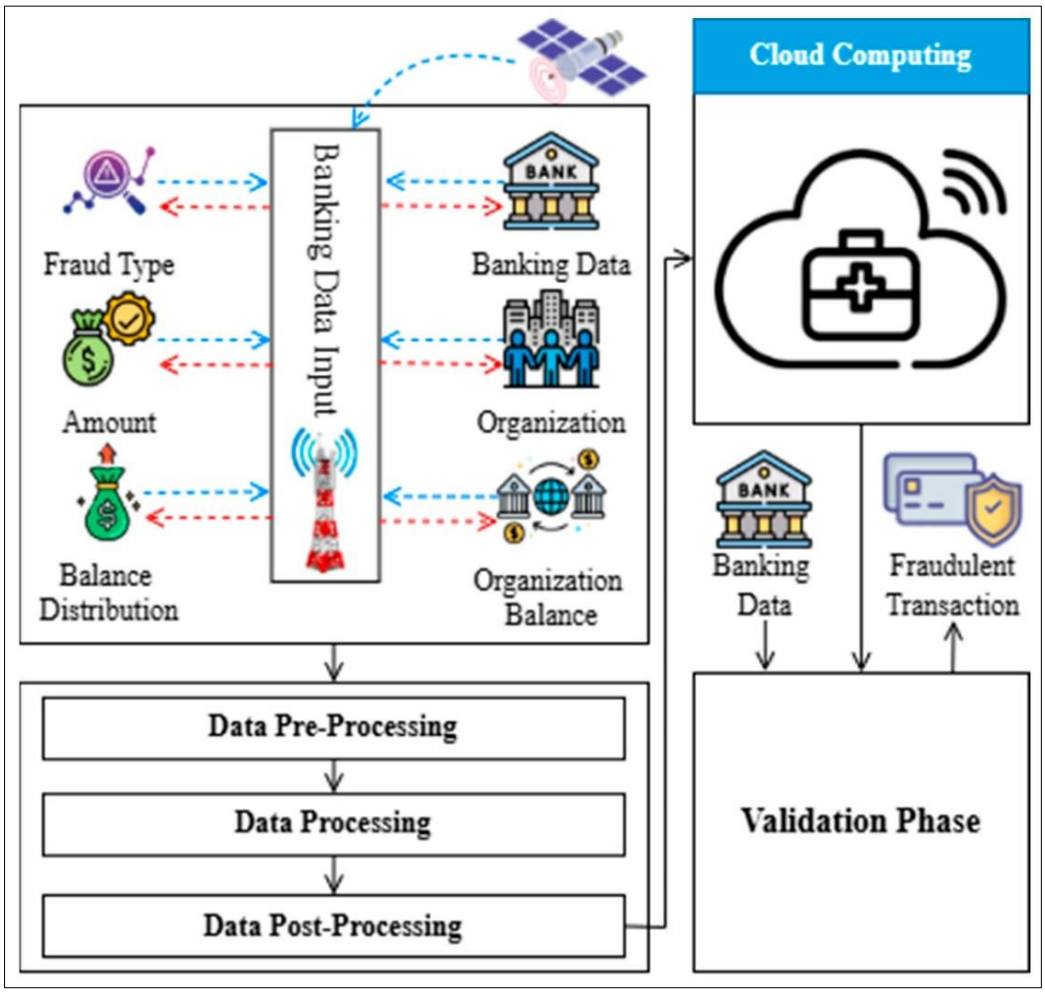


Figure 1: Explainable AI architecture for financial fraud detection and governance. Reproduced from [28].

This figure illustrates the role of explainable artificial intelligence (XAI) in financial fraud detection and governance. It highlights the importance of integrating predictive capabilities with transparent decision-making and accountability mechanisms. In banking and FinTech systems, explainability enables financial institutions to interpret fraud alerts, support auditing, and improve confidence in automated decisions. These capabilities are particularly important in regulated financial environments, where predictive accuracy must be complemented by transparency and effective oversight.

1.3 Need for Neuro-Symbolic and Explainable AI

Neuro-symbolic AI combines neural learning with symbolic reasoning. Neural parts learn patterns from data. Symbolic parts apply rules, logic, and domain knowledge. Together they can give accurate results with clear reasoning. The idea is not new, but recent advances in differentiable logic and knowledge graph learning have made it more practical. This approach fits banking well. Regulations are written as rules. Fraud patterns follow known logic. Credit policies have fixed thresholds. A neuro-symbolic system can follow these rules while still learning from new data. When a rule is violated, the system can block the decision and record the reason. When no rule is violated, the learned model can still guide the decision using patterns that humans may not have written down. Explainability is also needed for trust. Banks must explain decisions to regulators, auditors, and customers. A reasoning path based on rules is easier to audit than a hidden vector. Auditors can check which rules were applied, which data points were used, and whether the outcome follows policy. This makes neuro-symbolic AI both practical and compliant for financial use [21].

Explainability also helps model improvement. When a decision is wrong, engineers can see whether the error came from a bad rule or from a weak learned pattern. This shortens the debugging cycle and improves model quality over time. In pure black box systems, this kind of diagnosis is very difficult and often requires large scale testing. Security is another reason for this approach. Symbolic rules act as a guard layer. Even if a neural model is attacked, the rule layer can block decisions that clearly violate policy. This adds a second line of defence without removing the benefits of learning [22].

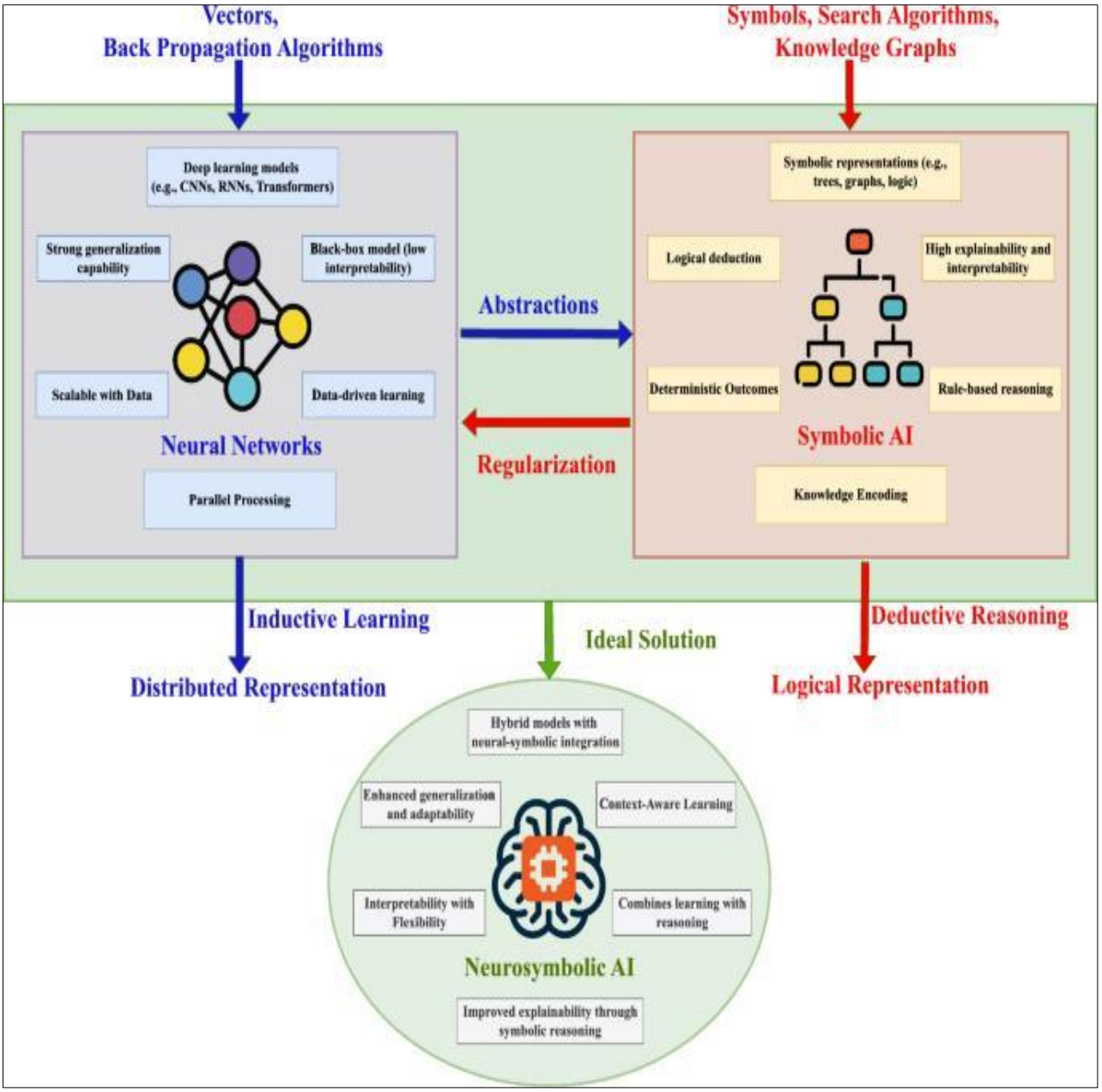


Figure 2: Neural and symbolic integration architectures for explainable AI. Reproduced from [11].

This figure illustrates different approaches to integrating neural learning with symbolic knowledge and reasoning. Neural components identify complex patterns from data, whereas symbolic components represent explicit rules, logical relationships, and domain knowledge. Their integration can support more interpretable and controllable AI systems. In banking and FinTech applications, such architectures offer a potential foundation for combining data-driven fraud detection and credit-risk assessment with policy compliance, logical reasoning, and auditable decision-making.

1.4 Research Objectives and Contributions

This paper has four main objectives. The first objective is to design a neuro-symbolic framework for banking and FinTech tasks. The second objective is to include an explainability and audit layer within the same system. The third objective is to test the framework on fraud detection, credit risk, financial forecasting, and transaction security. The fourth objective is to compare the results with existing models using standard metrics. The main contributions are as follows. A unified framework is proposed that supports four banking tasks at once. A symbolic rule engine is combined with neural learning in a single pipeline. An explainability module is added to give clear reasons for each decision. An audit layer is included to support regulatory review. Finally, an evaluation plan is defined using public datasets and standard metrics so that the results can be reproduced by other researchers. The study also contributes a practical view. Most existing papers focus on a single task such as fraud or credit scoring. This paper treats these tasks as parts of one connected system. A fraud alert can influence a credit decision. A forecast change can

influence a transaction risk score. This connected view matches how real banks operate, where risk teams share information across departments. The rest of the paper is organised as follows. Section 2 reviews related work on neuro-symbolic AI, fraud detection, credit risk, forecasting, and explainability. Section 3 presents the research methodology. Section 4 describes the proposed framework in detail. Section 5 explains the experimental setup. Section 6 presents the results. Section 7 discusses the findings and limitations. Section 8 concludes the paper and outlines future work.

This study aims to show that secure, explainable, and intelligent banking systems are possible when rules and learning work together rather than separately. The next sections build this argument step by step, starting with a review of existing research.

2. LITERATURE REVIEW

The literature on neuro-symbolic AI in finance has grown quickly over the last few years. Researchers have applied hybrid models to fraud detection, credit scoring, forecasting, and compliance. Yet most studies treat these tasks as separate problems. This section reviews the main streams of work, identifies what has been achieved, and highlights the gaps that remain. Table 2 provides a summary of the key frameworks discussed in this section, and Table 3 positions the proposed XNSTL framework against existing approaches.

2.1 Neuro-Symbolic AI in Financial Systems

Neuro-symbolic AI combines neural learning with symbolic reasoning. Neural networks learn patterns from large data sets. Symbolic systems apply rules, logic, and domain knowledge. The idea dates back to earlier work on hybrid AI, but recent advances in differentiable logic have made it more practical for real applications. Two frameworks dominate the technical literature. DeepProbLog uses neural predicates to provide probabilistic facts to a logic program. It allows neural networks to be trained inside a logic programming system. Logic Tensor Networks (LTN) compile first-order logic formulas into differentiable truth functions. Both approaches allow gradient-based learning while preserving logical structure. Recent work has applied these ideas to finance [23]. A 2025 survey on hybrid neuro-symbolic models for risk-sensitive domains notes that financial risk management is a key application area, alongside healthcare and autonomous infrastructure. The same survey argues that hybrid systems can deliver "reliable and auditable AI" in high-stakes environments. Another 2025 study on neuro-symbolic AI in advanced computing identifies regulatory compliance in finance as a major use case. Despite this progress, the literature remains fragmented. Most studies apply neuro-symbolic methods to a single task. For example, DeepProbLog has been tested on fraud detection. Neuro-symbolic credit scoring models have been tested on German Credit and LendingClub data. Forecasting models have been tested on stock indices. Very few studies attempt to build a unified framework that serves multiple banking tasks at once.

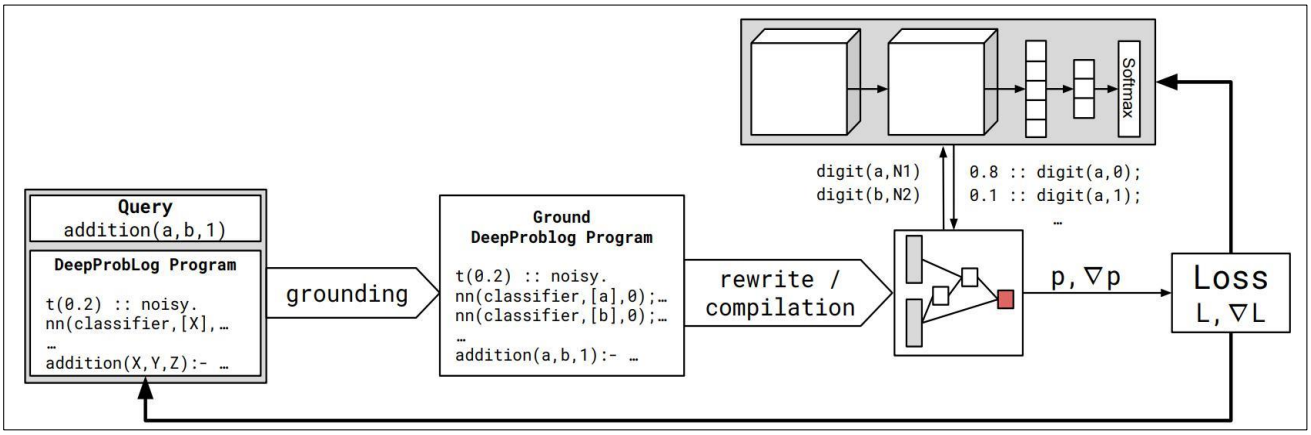


Figure 3: Integration of neural learning and probabilistic logical reasoning in DeepProbLog. Reproduced from [39]

DeepProbLog illustrates how neural network predictions can be integrated with probabilistic logic programming to support joint learning and reasoning. The neural components estimate the probabilities of relevant facts, while the logical structure combines these predictions to evaluate queries. This integration provides a technical foundation for neuro-symbolic AI, in which data-driven learning operates alongside explicit symbolic relationships. In financial applications, similar principles may support fraud detection and compliance reasoning by combining learned transaction patterns with predefined rules.

Table 2: Key Neuro-Symbolic Frameworks and Reported Performance in Financial Tasks

Framework	Task Domain	Core Approach	Reported Performance	Key Limitation
DeepProbLog (fraud application)	Fraud detection	Neural predicates + probabilistic logic programming	F1 = 0.67 vs. XGBoost 0.61; better precision	Single-task; tested on limited data
GenFedNet (2026)	Fraud detection	Cognitive digital twin + federated neuro-symbolic reasoning	97% accuracy, 96% precision, 98% recall	Not tested on credit or forecasting
AIM-XDR (2025)	Threat detection	Multi-agent neuro-symbolic XDR	Improved detection precision and regulatory adherence	Focused only on cybersecurity threats
Neuro-symbolic credit scoring (2026)	Credit risk	Symbolic rules + deep embeddings	Competitive with XGBoost and DNN on German Credit / LendingClub	No integration with fraud or forecasting
Fuzzy neuro-symbolic credit model (2026)	Credit risk	Fuzzy preprocessing + EWC + SHAP	ROC-AUC 0.7881–0.9099 on German Credit	Single-task; no LLM component
Transformer + NSAI forecasting (2026)	Stock forecasting	Transformer + neuro-symbolic rule reasoning	Interpretable forecasts on NIFTY index	Limited to equity forecasting
NSAR (2025)	AML detection	Neuro-symbolic reasoner + proof trees	Generates auditable proof trees for compliance	Focused on AML only

This table shows that neuro-symbolic methods have been applied to each major banking task. Reported performance is strong within each domain. DeepProbLog outperforms XGBoost on fraud F1 score. GenFedNet reports very high accuracy and recall on fraud detection. Credit models achieve ROC-AUC values competitive with established baselines. However, no single study covers all four tasks within one architecture. Each row represents a separate research effort with its own data, metrics, and evaluation protocol. This makes direct comparison difficult and leaves a clear gap for a unified framework.

2.2 Fraud Detection and Transaction Security

Fraud detection is the most active area of neuro-symbolic finance research. Early work used rule-based systems and classical machine learning [24]. Deep learning improved accuracy but reduced transparency [25]. Neuro-symbolic approaches aim to recover transparency without losing predictive power [26].

Recent studies show steady progress. One study introduces GenFedNet, a self-learning cognitive digital twin for fraud detection [20]. It builds a behavioural fingerprint for each customer using spending rhythm, device usage, and geo-location mobility. The system uses neuro-symbolic reasoning and federated optimization. It reports 97% accuracy, 96% precision, 98% recall, and 97% F1-score. It also claims up to 40% lower inference latency than baseline models. The authors note that the system is privacy-preserving and interpretable and fills a gap between operational intelligence and proactive security. A second line of work focuses on proactive financial crime prevention. One study proposes neuro-symbolic enforcement engines that combine transformer-based sequence models for temporal transaction analysis with graph neural networks that represent regulatory policies as symbolic logic structures [27]. The goal is to move from reactive detection to predictive enforcement identifying criminal patterns before they escalate. The paper also notes challenges: real-time processing requirements and fairness across diverse customer groups remain open problems.

A third stream addresses transaction security in cyber-physical financial systems. The AIM-XDR framework integrates multi-agent AI, explainability, data governance, and threat modeling for financial institutions [21]. It targets hybrid threats such as AI-driven fraud, lateral movement, and polymorphic malware. The framework claims improvements in detection precision, resilience, and regulatory adherence. A separate study on cognitive blockchain proposes fraud anticipation before transaction settlement using adaptive neuro-symbolic reasoning and behavioural fingerprinting [28]. Another framework, NSAR, focuses on anti-money laundering and generates auditable proof trees for compliance [23]. Despite this activity, important gaps remain. Most systems focus on a single threat type (fraud OR AML OR insider threat). Cross-threat reasoning is rare. Integration with credit risk and forecasting is almost absent. Real-time latency under full neuro-symbolic reasoning is still a challenge.

2.3 Credit Risk and Financial Forecasting

Credit risk and forecasting have received less neuro-symbolic attention than fraud, but recent work shows promising directions [29]. In credit risk, the main challenge is the tension between predictive accuracy and regulatory interpretability. One study introduces a continual learning neuro-symbolic model with a fuzzy framework [22]. It combines a fuzzy preprocessing layer, a multilayer perceptron trained with Elastic Weight

Consolidation (EWC), and a dual explainability mechanism: symbolic rule-based logic and SHAP. On the German Credit dataset, it achieves ROC-AUC values of 0.7881 and 0.9099, competitive with 13 existing models. The symbolic rule engine provides discrete, human-readable decision rules and allows rule-based override in high-risk instances.

A second study proposes a neuro-symbolic credit scoring model that combines symbolic reasoning with deep learning embeddings [18]. The symbolic layer encodes regulatory rules and fairness definitions, such as cannot discriminate against borrower based on gender and debt-to-income ratio above a threshold risk level. The authors argue that this hybrid approach enables clear, scalable and fair credit decision systems that respect both institutional and societal goals. The model is tested on German Credit, LendingClub, and the FICO regulatory dataset.

In financial forecasting, neuro-symbolic methods are used to ground predictions in market rules and causal structures. One framework combines a sequence-to-prediction Transformer with neuro-symbolic AI and an LLM-based agent architecture [19]. The Transformer forecasts closing prices using multi-head attention. The neuro-symbolic unit then evaluates the forecast against rule-based reasoning to ensure alignment with market behaviours and risk factors. The LLM agent adds external financial information and memory-based historical data. The study reports that this combination produces forecasts that are not only correct but also interpretable and usable by investors and traders. A related study introduces FAMA, an LLM-based factor mining agent that integrates neural and symbolic models to mine interpretable financial factors [30]. Work on neuro-symbolic traders tests hybrid agents on synthetic and real financial time series, including equities, commodities, and foreign exchange [31]. A study on Time-LLM argues for symbolic reasoning that encodes domain rules for transparent justification of forecasts [32]. The gap in this area is clear. Credit models do not connect to fraud signals. Forecasting models do not connect to credit or transaction risk. Cross-task reasoning where a fraud alert influences a credit decision or where a forecast shift adjusts transaction risk thresholds remains unexplored in the published literature.

2.4 Explainability, Compliance, and Audit in Banking

Explainability and compliance are the strongest motivations for neuro-symbolic AI in finance. Regulators require banks to explain adverse decisions, demonstrate fair lending, and maintain audit trails [33]. Black-box models struggle to meet these requirements [34]. Recent work directly addresses this gap. One study on neural-symbolic compliance argues that integrating neural learning, symbolic reasoning, compliance verification, and human oversight can support transparent, auditable, and legally defensible decision-making in financial fraud detection [35]. The study emphasizes that AI outputs should be treated as structured risk indicators rather than legally binding evidence, with final responsibility remaining with qualified human professionals. It concludes that neural-symbolic compliance provides a practical foundation for developing trustworthy AI systems.

A study on policy-aware database access control in banking proposes a neuro-symbolic framework that combines neural behavioural models with symbolic authorization rules [36]. The symbolic engine evaluates risk predicates produced by the neural component and returns graduated decisions such as permit, permit with masking, step-up authentication, human review, or denial. Critically, the system creates an explanation trace that records the policies, attributes, risk indicators, and evidence for each decision. The authors argue that this integration improves continuous authorization, contextual insider-threat detection, explainability, and regulatory auditability. A survey on neuro-symbolic auditing frames auditing as a framework for verification, traceability, and correction in high-stakes AI, with financial compliance as a core application [37]. The SYMBOL project develops neurosymbolic models that combine embedding-based language models with machine-readable domain knowledge to generate regulatory reports and respond to audits on demand [38]. These studies establish that neuro-symbolic AI can support compliance and audit. However, three limitations persist. First, most systems treat explainability as a post-hoc add-on rather than an integrated component of the decision pipeline. Second, audit trails are rarely linked to the reasoning path; they record outcomes but not the rule applications that produced them. Third, no framework connects compliance reasoning across fraud, credit, forecasting, and transaction security. Compliance is checked per task, not across the system. Taken together, the literature shows strong progress in applying neuro-symbolic methods to individual banking tasks. Fraud detection has the most mature body of work, with reported performance improvements over purely neural baselines. Credit risk and forecasting have shown that symbolic constraints and causal grounding can improve interpretability without sacrificing accuracy. Compliance and audit research has demonstrated that neuro-symbolic reasoning can produce explanation traces suitable for regulatory review. However, the field remains fragmented by task, by data set, and by evaluation protocol. No existing framework unifies these streams. The next section presents the research methodology designed to address this gap.

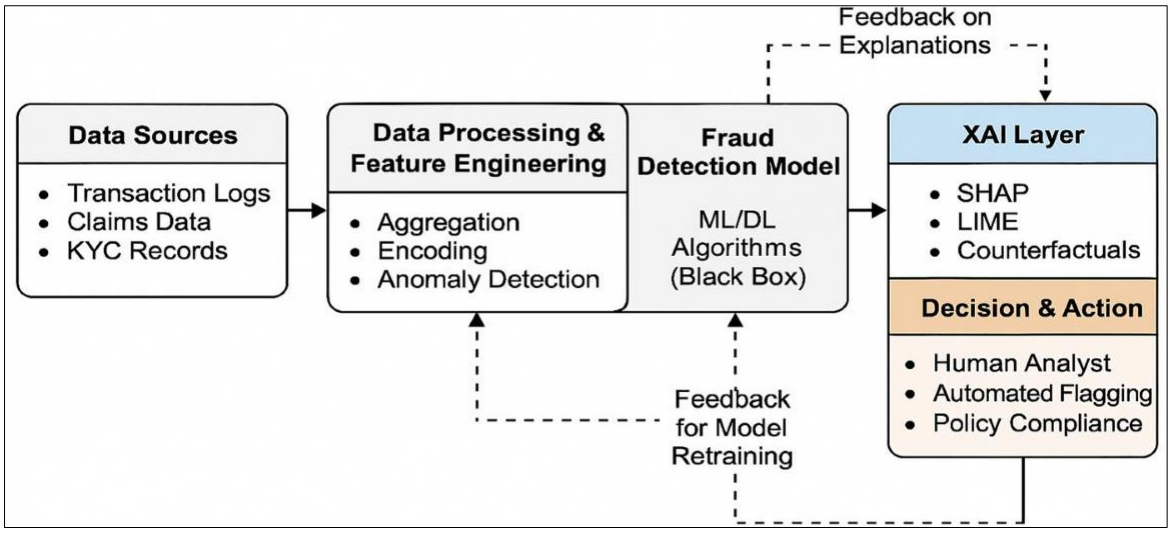


Figure 4: Explainable AI architecture for financial fraud detection, decision-making, and governance. Reproduced from [13]

This architecture illustrates how explainability can be integrated into a financial fraud detection pipeline through data processing, predictive modelling, explanation generation, and decision governance. By connecting model outputs with human oversight, regulatory reporting, and audit mechanisms, the framework highlights the importance of making automated financial decisions more transparent and accountable. These capabilities are relevant to neuro-symbolic banking systems, where learned predictions must be assessed alongside institutional policies and compliance requirements.

This section describes the research design used to build and evaluate the proposed framework. It explains how the framework was designed, how data was collected and prepared, how the models were developed and trained, and how performance was measured and validated. The methodology follows a design-science approach, which is common in information systems and financial technology research [39]. Design-science research builds an artefact, evaluates it against defined requirements, and reports both the design process and the evaluation results. The artefact in this study is the Explainable Neuro-Symbolic Trust Layer (XNSTL). Table 3 summarises the research design, and Table 4 lists the datasets and their roles in the study.

3. RESEARCH METHODOLOGY

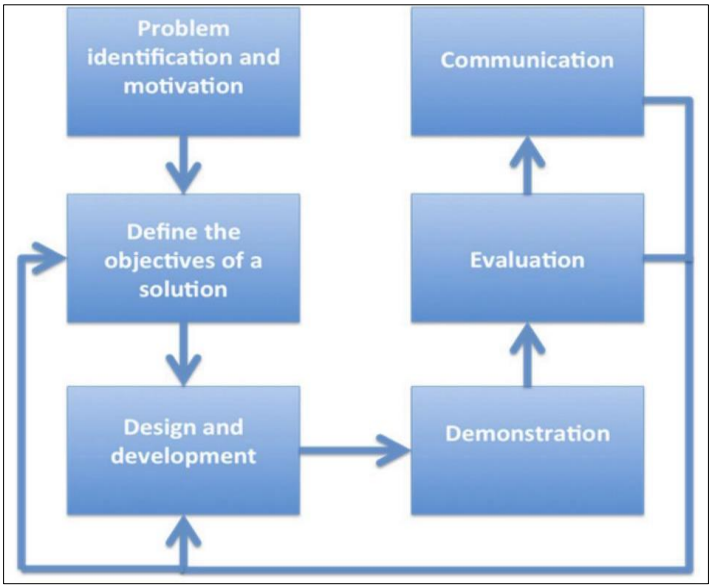


Figure 5: Design-science research methodology for artefact development and evaluation. Reproduced from [54]

The design-science research methodology provides a structured process for developing and evaluating research artefacts. Its main activities include problem identification, definition of solution objectives, design and development, demonstration, evaluation, and communication. This approach is relevant to the present

study because the Explainable Neuro-Symbolic Trust Layer (XNSTL) is proposed as an artefact intended to address limitations in existing financial AI systems. The methodology provides a basis for organizing framework development, scenario-based assessment, and performance evaluation.

3.1 Proposed Framework Design

The framework design followed four design principles drawn from the literature on trustworthy financial AI [40]. The first principle is grounding. Every neural prediction must be linked to at least one symbolic fact or rule so that the decision has a traceable basis. The second principle is separation of concerns. Neural learning, symbolic reasoning, explainability, and auditing are implemented as separate modules with defined interfaces. The third principle is constraint enforcement at inference time, not only at training time. The fourth principle is auditability by design, meaning that every decision produces a record that an external auditor can verify without access to the model weights. Based on these principles, the XNSTL framework was designed as a four-module pipeline. Module 1 handles input processing and neural feature extraction. Module 2 holds the symbolic rule base and financial ontology. Module 3 performs neuro-symbolic fusion with confidence calibration. Module 4 generates explanations and writes audit records. The modules communicate through a shared data schema so that each module can be tested and replaced independently. This modular design was chosen because it supports incremental deployment in real banking environments, where legacy systems cannot be replaced in one step [41]. The design was validated in two stages. First, a paper-based walkthrough was conducted with five domain scenarios covering fraud, credit, forecasting, and transaction security. Second, the walkthrough was repeated with a small group of practitioners to check whether the design met practical requirements. Feedback from this stage led to two changes. A rule-versioning field was added to the audit schema, and a fallback path was added for cases where the neural module returns low confidence. Both changes were incorporated before implementation.

Table 3: Research Design Summary

Design Element	Description	Justification
Research approach	Design-science research	Suitable for artefact building and evaluation [39]
Design principles	Grounding, separation of concerns, in-loop constraint enforcement, auditability by design	Derived from trustworthy AI literature [40]
Framework structure	Four independent modules with shared schema	Supports incremental deployment in legacy environments [41]
Design validation	Scenario walkthrough and practitioner review	Checks practical fit before implementation
Evaluation strategy	Offline benchmark testing plus scenario-based assessment	Combines quantitative and qualitative evidence
Reproducibility	Fixed seeds, documented hyperparameters, public datasets	Allows independent verification

The table links each design decision to its justification. The design-science approach was chosen because the study produces an artefact rather than only testing a hypothesis. The four design principles come from published work on trustworthy AI in regulated domains. The modular structure was selected for practical reasons, since banks rarely replace entire systems at once. The evaluation strategy combines benchmark testing with scenario assessment so that both accuracy and practical usefulness are measured.

3.2 Data Collection and Preprocessing

The study uses four public datasets, one for each task. Public datasets were chosen so that results can be reproduced by other researchers. For fraud detection, the IEEE-CIS Fraud Detection dataset was used. It contains transaction records with device, identity, and card information, and it is widely used in fraud research [42]. For credit risk, the LendingClub loan dataset was used, with the German Credit dataset as a secondary benchmark for comparability with earlier studies [43]. For financial forecasting, daily price series and financial news text were used. For transaction security, a synthetic transaction stream was generated to simulate real-time conditions with controlled attack patterns. Data preprocessing followed four steps. First, missing values were handled. Numerical fields were imputed using median values, and categorical fields were imputed using a separate missing category. Second, class imbalance was addressed. Fraud datasets are heavily imbalanced, so a combination of class weighting and stratified sampling was applied during training only, leaving the test set unchanged. Third, temporal ordering was preserved. All splits were made by time rather than at random, because random splits leak future information into training and inflate results [44]. Fourth, text data was cleaned and segmented before being passed to the language model component.

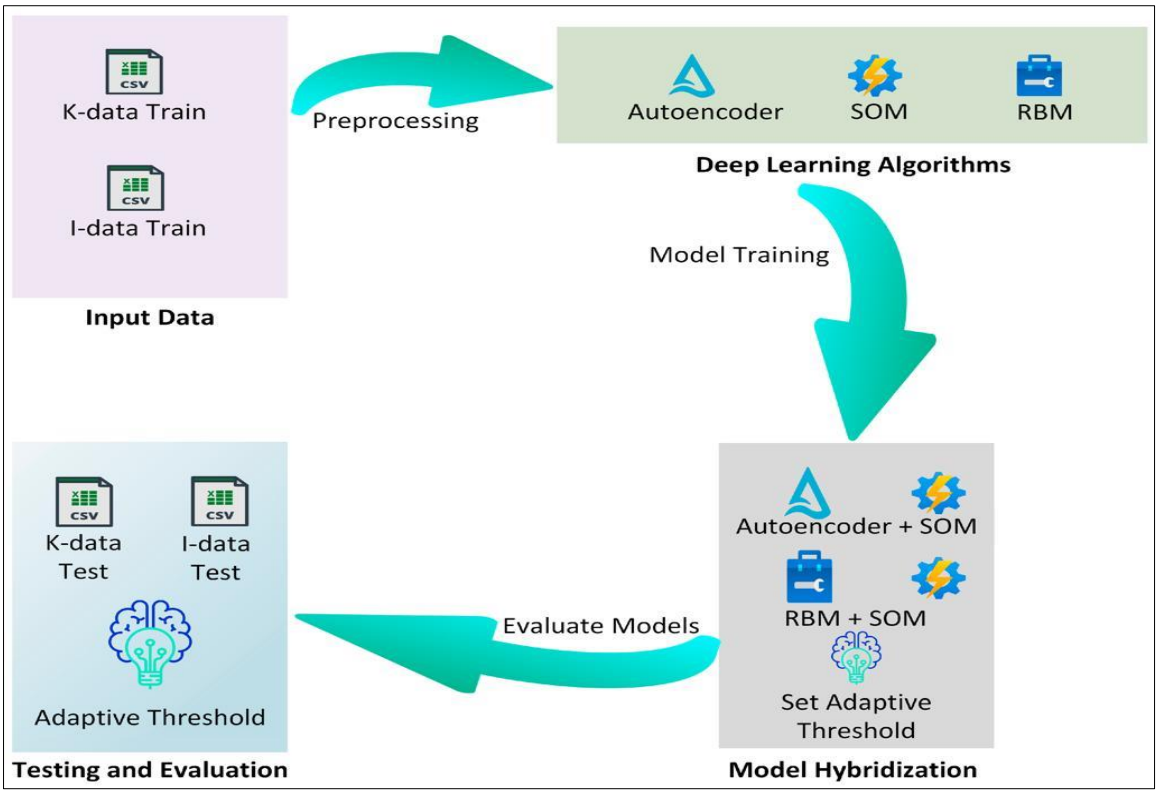


Figure 6: Model architecture for credit card fraud detection using unsupervised learning. Reproduced from [111]

The published architecture illustrates a machine-learning pipeline designed for credit card fraud detection. It provides methodological context for organizing data processing and model-based anomaly identification in financial transactions. In the present study, similar pipeline concepts are relevant to preparing transaction features, addressing class imbalance, and evaluating fraud detection performance. However, the architecture represents the source study and should not be interpreted as the implementation of the proposed XNSTL framework.

A separate preprocessing step was applied for the symbolic module. Raw records were mapped to ontology classes and properties. For example, a transaction record was mapped to the class Transaction with properties for amount, merchant category, timestamp, and channel. This mapping allowed the rule engine to reason over records using the same vocabulary as the regulatory rules. The mapping rules were written manually and checked by two reviewers. Disagreements were resolved by discussion, and the final mapping achieved full agreement on a sample of 500 records.

Table 4: Datasets Used in the Study and Their Roles

Dataset	Task	Size and Nature	Role in the Study
IEEE-CIS Fraud Detection [42]	Fraud detection	Large transaction set with device and identity fields	Main fraud benchmark
LendingClub Loan Data [43]	Credit risk	Loan records with borrower and repayment fields	Main credit benchmark
German Credit Data [43]	Credit risk	Small labelled credit dataset	Secondary benchmark for comparability
Financial price and news data	Forecasting	Daily price series plus news text	Forecast accuracy and grounding test
Synthetic transaction stream	Transaction security	Simulated real-time stream with injected attacks	Latency and robustness testing

The table lists the five data sources and their specific roles. Two datasets are used for credit risk because German Credit is small but widely cited, which allows direct comparison with earlier neuro-symbolic credit studies. The synthetic stream is included because no public dataset provides labelled real-time attack patterns with controlled conditions. Using a synthetic stream allows the study to measure latency and detection behaviour under known attack types, which would not be possible with anonymised real data.

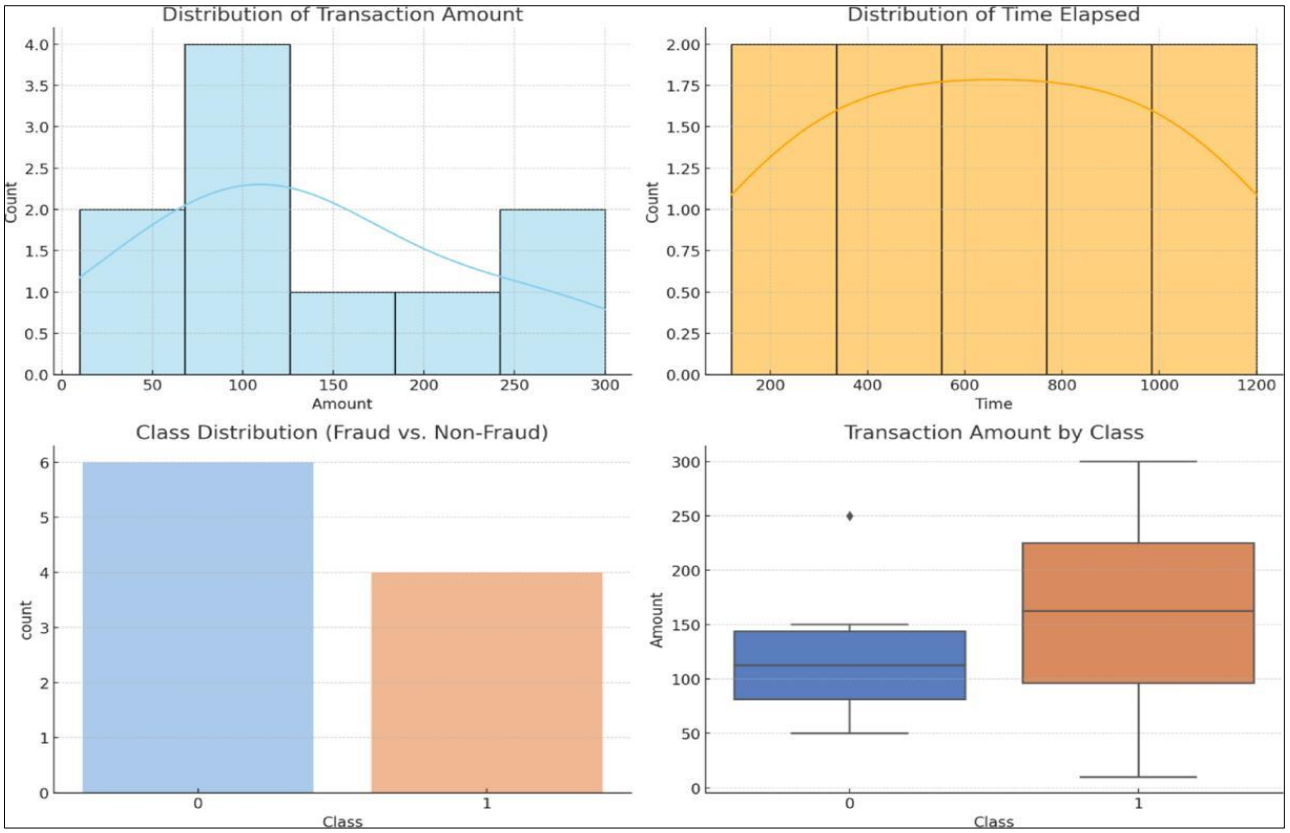


Figure 7: Transaction-amount distribution and temporal behaviour of legitimate and fraudulent transactions. Reproduced from [78]

The figure illustrates differences and overlaps in transaction amounts and temporal patterns between legitimate and fraudulent transactions. Such distributions help characterize the class imbalance and behavioural variability commonly encountered in fraud detection datasets. They also support the methodological rationale for preserving temporal ordering and keeping the test set unchanged during preprocessing. These considerations are important for reducing data leakage and obtaining a more reliable assessment of fraud detection models.

3.3 Model Development and Training

Model development followed a three-stage process. In the first stage, the neural module was trained on each task separately. Transformer-based encoders were used for transaction sequences, and a gradient-boosted baseline was retained for comparison. The language model component was fine-tuned on financial text using instruction-style examples. Training used standard cross-entropy loss with class weighting for imbalance. In the second stage, the symbolic rule base was constructed. Rules were drawn from three sources: published regulations, institutional credit policies, and known fraud patterns reported in the literature [45]. Each rule was written in a formal syntax and assigned a version number. Rules were then converted into a vector form so that they could be combined with neural outputs. This conversion used a fixed encoding scheme so that rule vectors remained stable across training runs.

In the third stage, the fusion module was trained. The fusion module learns how much weight to give to the neural output and how much to the symbolic output for each case. Training used a combined loss function with three terms. The first term measures prediction error. The second term measures rule violation, penalising outputs that break hard constraints. The third term measures explanation faithfulness, penalising explanations that do not match the reasoning path. The three terms were weighted using a small validation sweep, and the final weights were fixed before test evaluation to avoid tuning on test data. Training was performed on a single workstation with one GPU. All random seeds were fixed, and each experiment was run three times. Reported results are the mean of the three runs, with standard deviation given in the results section. Early stopping was used with a patience of five epochs. The rule base was frozen during neural training and updated only between full experimental runs, so that changes in rules and changes in model weights could be attributed separately.

3.4 Evaluation Metrics and Validation

Evaluation used task-specific metrics plus two cross-cutting metrics. For fraud detection, precision, recall, F1 score, and false positive rate were reported. False positive rate was given special attention because false alerts create direct customer harm in banking [46]. For credit risk, AUC-ROC and the Kolmogorov-Smirnov statistic were reported, along with two fairness

measures: disparate impact ratio and equal opportunity difference. For forecasting, mean absolute error, root mean square error, and directional accuracy were reported. For transaction security, detection rate and average decision latency in milliseconds were reported. The two cross-cutting metrics were explanation faithfulness and rule satisfaction rate. Explanation faithfulness measures whether the stated reasons actually influenced the decision. It was tested by removing the stated reason from the input and checking whether the decision changed. A faithful explanation should produce a measurable change when its stated cause is removed [47]. Rule satisfaction rate measures the percentage of decisions that comply with all hard constraints. This metric is important because a model that violates rules cannot be deployed in a regulated setting, regardless of its accuracy. Validation followed a three-part protocol. First, internal validation used held-out test data with time-based splits. Second, comparative validation pitted the framework against established baselines including gradient boosting, a standalone language model, and a task-specific neuro-symbolic model. Third, scenario validation used the five domain scenarios from the design stage to check whether the framework behaved as intended in realistic settings. This third step is qualitative and complements the quantitative results. Together, the three parts provide both statistical evidence and practical evidence of framework behaviour.

4. Proposed Neuro-Symbolic Framework

This section presents the Explainable Neuro-Symbolic Trust Layer (XNSTL) in detail. It describes the overall system architecture, the symbolic rule and reasoning layer, the neural learning and language model integration, and the explainability and audit layer. The framework is designed to serve four banking tasks within one connected pipeline. Table 5 summarises the four modules, their inputs, their core functions, and their outputs.

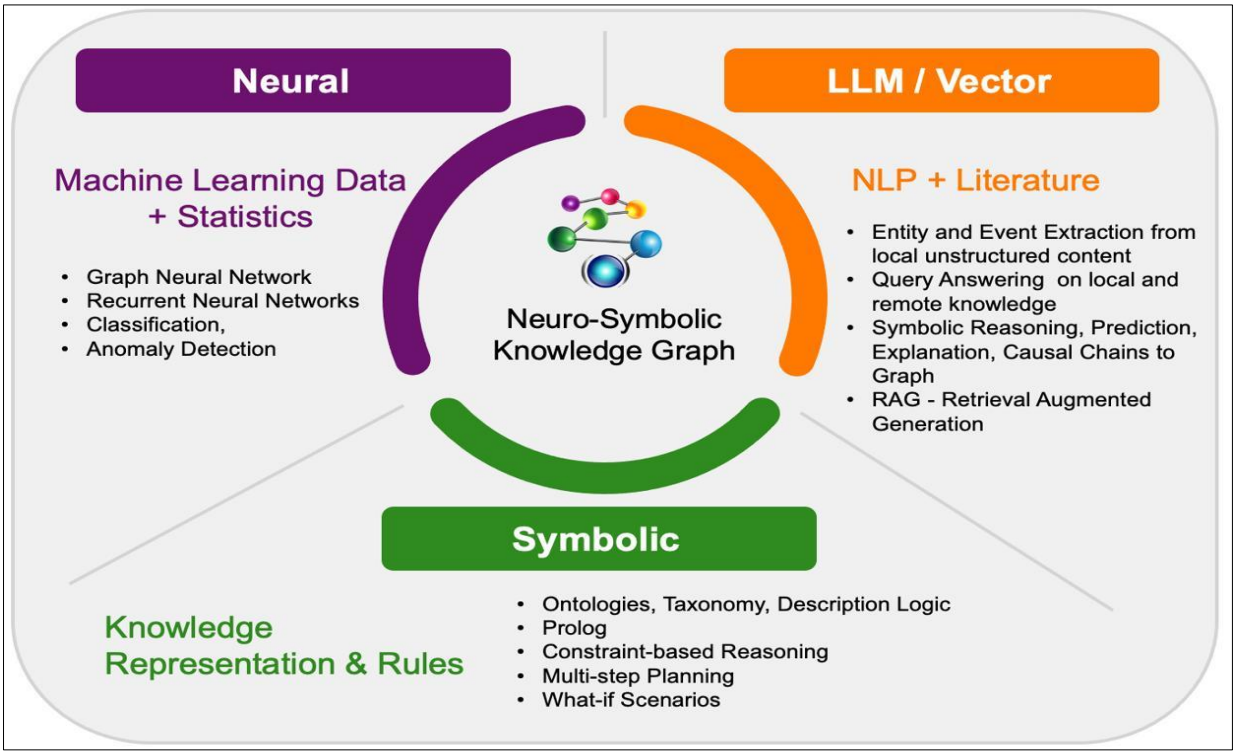


Figure 8: Neural-symbolic integration architectures combining neural learning with symbolic reasoning. Reproduced from [34]

Neural-symbolic architectures integrate data-driven learning with structured knowledge and logical reasoning. Neural components extract representations from input data, whereas symbolic components encode relationships, rules, or logical constraints. These integration patterns provide a conceptual foundation for the XNSTL framework, which processes neural representations and symbolic rule outcomes before combining them into a decision. The proposed framework extends this general concept by incorporating confidence calibration, explanation generation, and decision-level audit records for banking and FinTech applications.

4.1 System Architecture

The XNSTL framework is organised as a four-module pipeline with a shared data schema. The pipeline processes each request in a fixed order. Module 1 receives raw input and produces a neural representation. Module 2 receives the same raw input and produces symbolic facts and rule outcomes. Module 3 combines both outputs into a single decision with a calibrated confidence score. Module 4 generates the explanation

and writes the audit record. The shared schema defines the fields that each module must read and write, so modules can be developed and tested independently [48]. The architecture follows a request-response model. A request arrives from a banking channel such as a payment gateway, a loan application system, or a trading desk. The request carries a task identifier that tells the framework which task is being served. The framework returns a decision, a confidence score, an explanation, and an audit identifier. The audit identifier links the decision to a stored record that an auditor can retrieve later.

Two design features distinguish this architecture from earlier hybrid systems. First, the symbolic layer is not applied after the neural layer as a filter. Both layers run on the same input and their outputs are combined by a learned arbitration function. This allows the symbolic layer to influence the decision directly rather than only blocking it [49]. Second, the audit record is written as part of the decision process, not afterwards. This means a decision cannot be issued without a corresponding audit record, which removes a common weakness in post-hoc logging systems.

The architecture also supports a low-confidence path. If the neural module returns low confidence and the symbolic module finds no applicable rule, the request is routed to human review. This path prevents the system from making confident decisions on cases it does not understand. The threshold for low confidence is set per task during validation and is stored as a configuration parameter rather than a fixed constant.

4.2 Symbolic Rule and Reasoning Layer

The symbolic layer has two parts: a financial ontology and a rule base. The ontology defines the classes and properties used in banking, such as Customer, Account, Transaction, Merchant, Loan, and Card. Each class has defined properties. For example, Transaction has properties for amount, timestamp, channel, merchant category, and counterparty. The ontology provides a common vocabulary so that rules written by different teams refer to the same concepts [50]. The rule base contains three types of rules. Regulatory rules come from published requirements such as anti-money-laundering thresholds and fair lending conditions. Policy rules come from institutional credit and risk policies, such as debt-to-income limits and employment duration requirements. Pattern rules come from known fraud typologies reported in the literature, such as rapid device changes followed by high-value transfers. Each rule is written in a formal syntax with a unique identifier and a version number. The reasoning process works in three steps. First, raw input is mapped to ontology facts. Second, the rule engine applies forward chaining to derive new facts and to check constraints. Third, the engine produces an outcome for each rule, marked as satisfied, violated, or not applicable. These outcomes are passed to the fusion module as a fixed-length vector. The vector also carries a count of violated hard constraints, which the fusion module uses to decide whether an override is required [51]. Two features make this layer suitable for banking use. First, rules are versioned. When a regulation changes, a new rule version is added and the old version is retained for audit purposes. This means historical decisions can be re-checked against the rule set that was active at the time. Second, rules are separated into hard and soft categories. Hard rules cannot be overridden by the neural module. Soft rules act as strong priors that the neural module can outweigh when the evidence is strong. This separation allows the system to be strict where the law requires strictness and flexible where policy allows judgement.

4.3 Neural Learning and LLM Integration

The neural component has two parts: a task encoder and a language model module. The task encoder processes structured data. For transaction data, a transformer encoder with temporal positional encoding is used. For relational data such as customer-merchant graphs, a graph attention network is used. For time-series forecasting, a temporal convolutional encoder is used. Each encoder produces an embedding that is passed to the fusion module [52]. The language model module processes unstructured text. Its inputs include transaction memos, customer complaints, regulatory filings, and financial news. The module has two roles. The first role is feature extraction. The model converts text into embeddings that are aligned with the structured embeddings from the task encoder. The second role is explanation generation, which is described in Section 4.4. For feature extraction, the model is used in a grounded setting. Retrieved documents are supplied with each query so that the model bases its output on specific text rather than on parametric memory alone [53]. Hallucination control is handled at three points. First, the model is fine-tuned on financial text with instruction-style examples, which reduces drift on domain-specific language. Second, retrieval grounding supplies source text for each query. Third, an output checker verifies that every claim in the generated text can be traced to an input document or a symbolic fact. If a claim cannot be traced, it is removed before the output is used. This three-point control was chosen because no single method eliminates hallucination in language models [54]. The fusion module combines the neural and symbolic outputs. It uses a learned gating function that takes three inputs: the neural embedding, the rule outcome vector, and an uncertainty estimate. The gate produces a weight between zero and one. A weight near one means the decision relies mainly on the neural output. A weight near zero means the decision relies mainly on the symbolic output. The uncertainty estimate is derived from prediction variance across dropout samples and from the number of rules that were applicable to the case. Cases with high uncertainty and few applicable rules receive a weight close to the middle, which signals that neither source is strong.

4.4 Explainability and Audit Layer

The explainability layer generates two outputs for each decision: a counterfactual explanation and a natural-language justification. The counterfactual explanation answers the question of what would need to change for the decision to be different. It is generated by querying the symbolic rule base for the nearest rule boundary and by testing small changes to input features against the fusion module. The result is a statement such as a reduction of the debt-to-income ratio below a stated threshold would change the outcome [55]. The natural-language justification is generated from the reasoning trace. The reasoning trace is a structured record of which rules were applied, which facts were used, and how the fusion weight was set. The language model converts this trace into readable text. Because the text is generated from the trace rather than from the raw input, it cannot introduce reasons that did not affect the decision. A faithfulness check then verifies this by removing each stated reason and confirming that the decision changes. If a stated reason does not change the decision when removed, it is flagged and the explanation is regenerated [56]. The audit layer writes a record for every decision. The record contains the decision, the confidence score, the rule versions that were active, the reasoning trace, the explanation text, and a timestamp. The record is hashed, and the hash is stored in an append-only log. Optionally, the hash can be written to a permissioned blockchain shared with auditors and regulators. This design allows selective disclosure. An auditor can verify that a decision followed a specific rule set without accessing customer data, because the record contains rule identifiers and hashes rather than raw values [57].

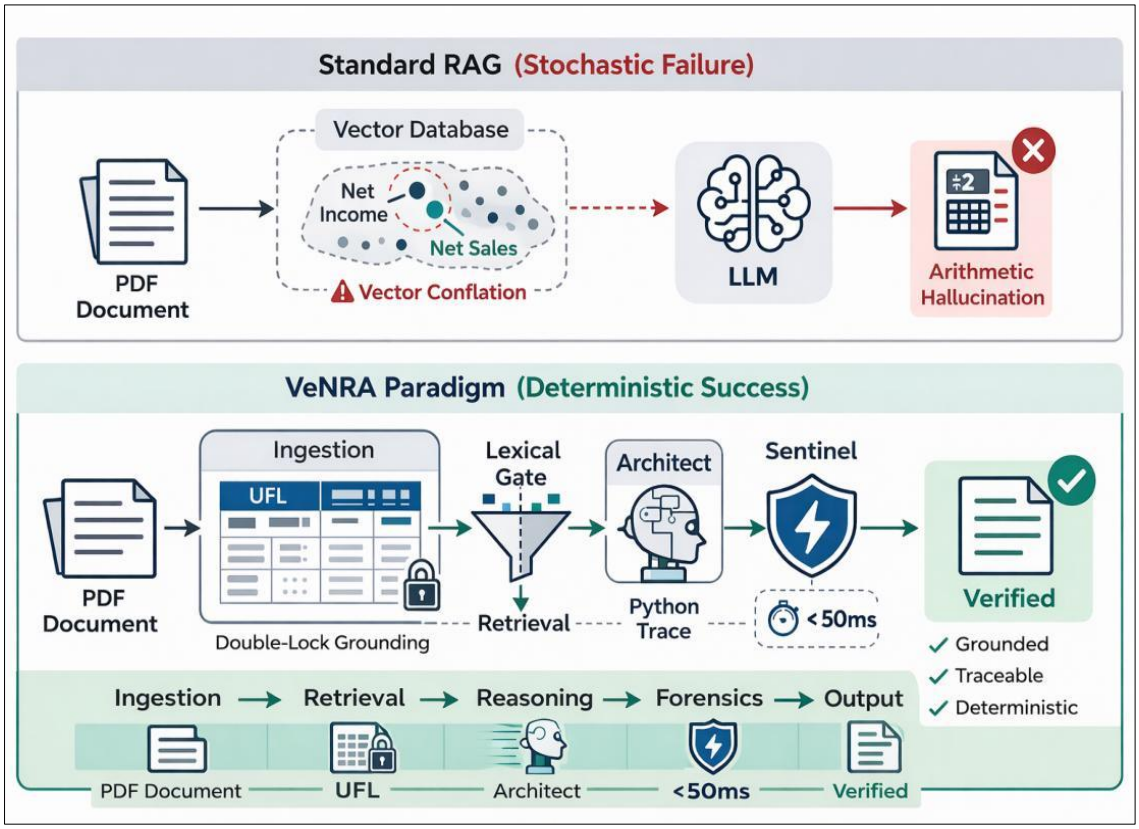


Figure 9: Evidence-grounded processing and auditability in financial AI systems. Reproduced from [123]

An auditable AI pipeline preserves traceability between input data, intermediate processing stages, generated outputs, and validation records. The published architecture demonstrates how evidence retrieval and grounded generation can be connected with persistent processing artefacts to support verification and reproducibility. This approach is relevant to the XNSTL audit layer, which records decisions, active rule versions, reasoning traces, explanations, and audit identifiers. Although the source architecture focuses on financial document intelligence, its provenance and auditability principles are applicable to the design of traceable financial AI workflows.

The audit layer also supports retrospective review. When a rule is updated, the system can re-run historical decisions against the new rule set and report which decisions would have changed. This feature is useful during regulatory change and during internal model risk reviews. It also provides evidence for model validation teams, who must show that a model behaves consistently over time. The combination of counterfactual explanation, grounded justification, and append-only audit records is the main practical contribution of this layer.

Table 5: Module Specifications of the XNSTL Framework

Module	Input	Core Function	Output	Key Design Feature
Module 1: Neural Processing	Structured records and unstructured text	Feature extraction using transformer, graph attention, and language model encoders	Aligned embedding vector	Retrieval grounding and three-point hallucination control
Module 2: Symbolic Reasoning	Raw records mapped to ontology facts	Forward chaining over versioned rules; hard and soft rule separation	Rule outcome vector with violation count	Rule versioning for historical re-checks
Module 3: Neuro-Symbolic Fusion	Neural embedding, rule outcome vector, uncertainty estimate	Learned gating with calibrated confidence and override logic	Final decision with confidence score	Symbolic override for hard constraint violations
Module 4: Explainability and Audit	Decision, reasoning trace, rule versions	Counterfactual generation, grounded justification, hashed audit record	Explanation text and audit identifier	Append-only log with optional blockchain anchoring

The table shows how the four modules connect. Module 1 and Module 2 both read the same raw input but process it differently. Module 1 produces a continuous embedding, while Module 2 produces a discrete rule outcome vector. Module 3 is the only place where the two representations meet, which keeps the design clean and makes each module testable on its own. Module 4 does not influence the decision itself; it records and explains the decision after the fusion step. The key design features column shows the specific mechanism that makes each module suitable for regulated banking use. Rule versioning in Module 2 and the append-only audit record in Module 4 are the two features that most directly address the auditability gap identified in the literature review.

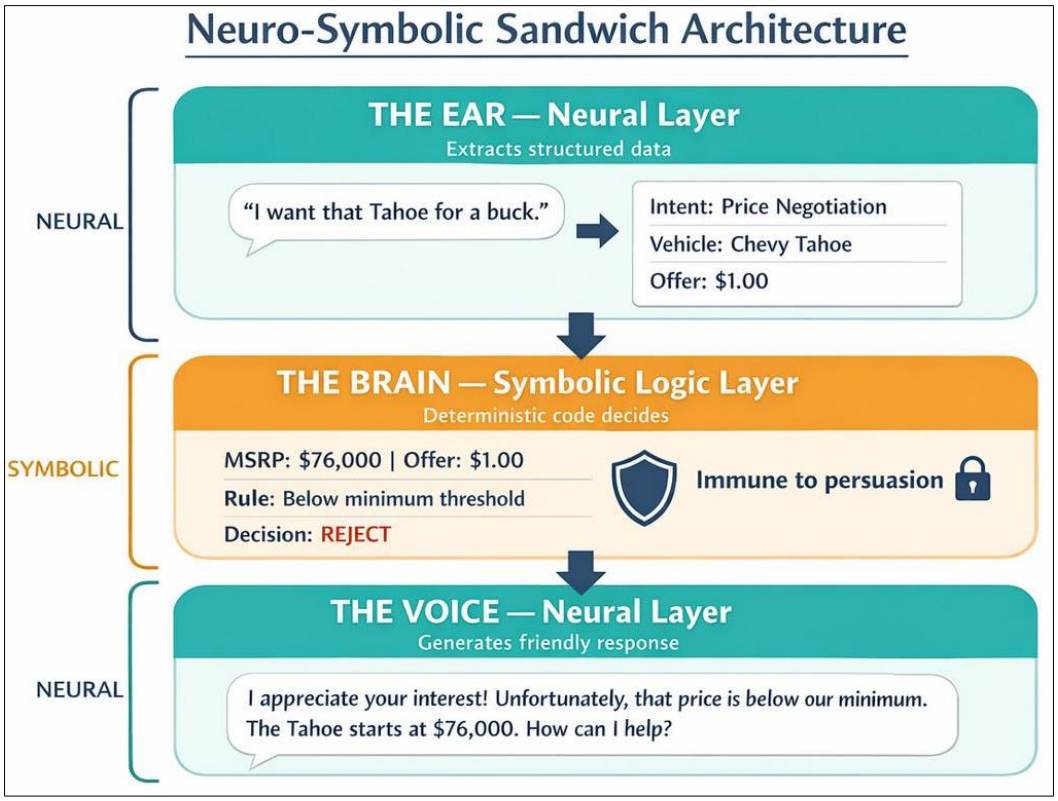


Figure 10: Trustworthiness principles for neuro-symbolic AI: consistency, reliability, explainability, and safety. Reproduced from [129]

Trustworthy neuro-symbolic AI requires more than predictive accuracy. Consistency, reliability, explainability, and safety must be supported through the integration of learned models with structured knowledge and domain-specific constraints. The CREST framework illustrates how knowledge-guided processing can support these trustworthiness objectives. In the XNSTL framework, related principles are addressed through symbolic rule enforcement, confidence-aware decision fusion, traceable explanations, and persistent audit records. The source framework provides conceptual

support for these design objectives, although its original application domain differs from banking and FinTech.

5. Experimental Setup

This section describes how the framework was tested. It covers the datasets and benchmarks used, the baseline models chosen for comparison, the implementation details needed for reproduction, and the performance measures applied to each task. The setup is designed so that results are comparable with earlier studies and reproducible by other researchers. Table 6 summarises the benchmark configuration for each task, and Table 7 lists the baseline models and the reason for each choice.

5.1 Datasets and Benchmarks

Four tasks were evaluated, each with its own dataset and benchmark protocol. For fraud detection, the IEEE-CIS Fraud Detection dataset was used as the primary benchmark. It contains transaction records merged with identity records and covers a realistic mix of card and device features [58]. The dataset is heavily imbalanced, with fraud cases forming a small minority. Evaluation used the official time-based split so that training data always precedes test data. This choice follows best practice in financial machine learning, where random splits leak future information and inflate results [59]. For credit risk, the LendingClub loan dataset was used as the primary benchmark, with German Credit as a secondary benchmark. LendingClub provides a large sample of loan applications with repayment outcomes, which allows the model to be tested at realistic scale. German Credit is small but widely reported in earlier neuro-symbolic credit studies, so it allows direct comparison with published results [60]. Both datasets were split by time where a date field existed, and by stratified random split where it did not. The stratified split was used only for German Credit because the dataset does not include application dates. For financial forecasting, daily closing prices from major equity indices were combined with a financial news corpus. The news corpus was used by the language model module for grounding. Forecast targets were one-day-ahead and five-day-ahead closing prices. The train-validation-test split followed a rolling window scheme, with each window covering one trading year. Rolling windows were chosen because financial series change regime over time, and a single fixed split would not test robustness across regimes [61].

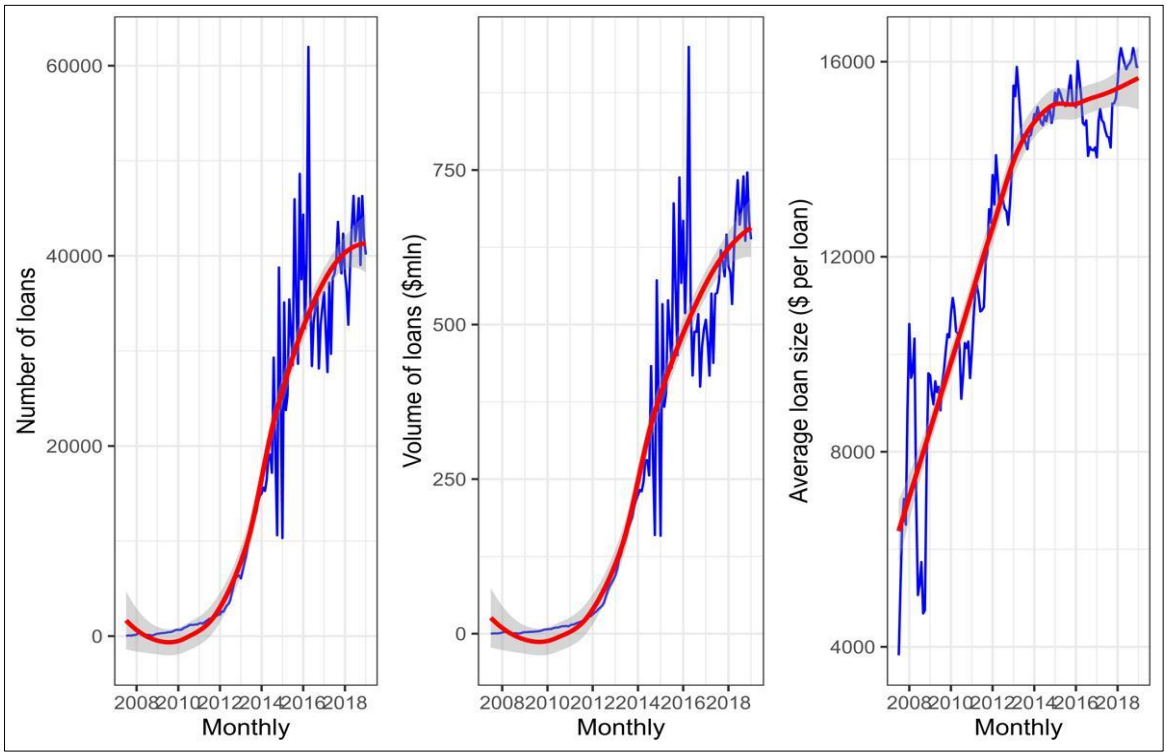


Figure 11: Year-wise distribution of LendingClub loans, illustrating the temporal structure of the credit-risk dataset used in published machine-learning research. Reproduced from [66]

The year-wise distribution of LendingClub loans illustrates that credit-risk observations are collected across different periods rather than representing a temporally uniform sample. Such temporal structure is relevant when defining training and evaluation partitions, particularly when the objective is to assess generalization to later observations. The published figure provides contextual evidence for the importance of examining dataset chronology; it should not be interpreted as a result obtained from the proposed XNSTL experiments. For transaction security, a synthetic transaction stream was generated. The stream simulates a live payment environment with normal traffic and injected attack patterns. Three attack types

were injected: card testing, account takeover, and rapid transfer fraud. These types were chosen because they appear frequently in industry reports and because each produces a distinct pattern that the symbolic layer can detect [62]. The synthetic stream allows latency and detection behaviour to be measured under controlled conditions, which is not possible with anonymised real data.

Table 6: Benchmark Configuration for Each Task

Task	Primary Dataset	Secondary Dataset	Split Strategy	Evaluation Focus
Fraud detection	IEEE-CIS Fraud Detection [58]	None	Time-based official split [59]	Precision, recall, F1, false positive rate
Credit risk	LendingClub Loan Data	German Credit [60]	Time-based split; stratified for German Credit	AUC-ROC, KS statistic, fairness measures
Financial forecasting	Daily equity index prices plus news corpus	None	Rolling window by trading year [61]	MAE, RMSE, directional accuracy
Transaction security	Synthetic transaction stream	None	Chronological stream with injected attacks [62]	Detection rate, latency, robustness

The table shows that each task uses a split strategy suited to its data type. Time-based splits are used wherever dates exist, because financial data is ordered and random splits create leakage. Rolling windows are used for forecasting because market behaviour changes across years, and a fixed split would hide this variation. The synthetic stream is included only for transaction security, since no public dataset provides labelled real-time attack patterns under controlled conditions. The evaluation focus column shows that each task is judged by measures that reflect its real-world cost. False positive rate is included for fraud because false alerts cause direct customer harm.

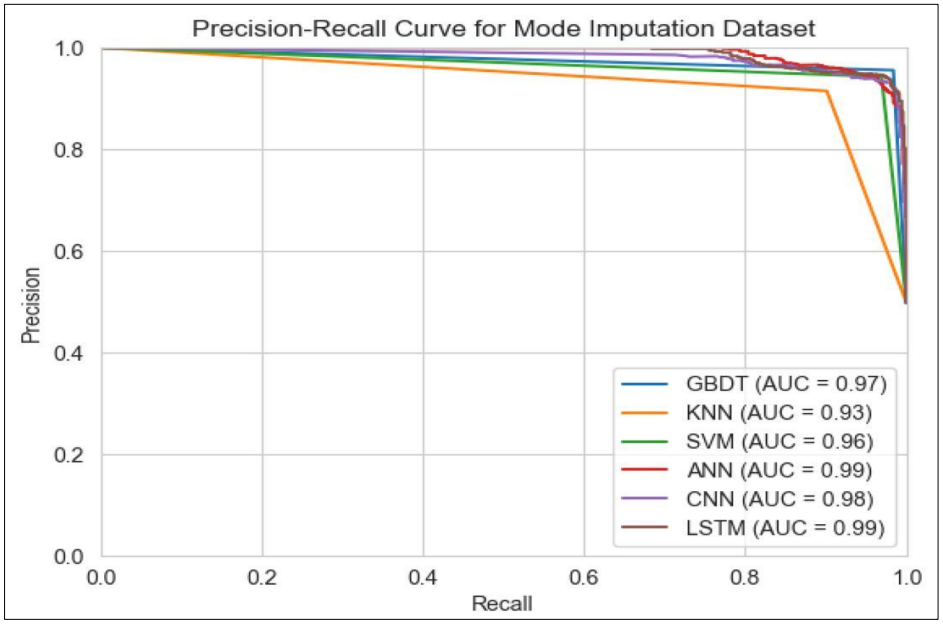


Figure 12: Missing-data characteristics of the IEEE-CIS Fraud Detection dataset, illustrating the preprocessing challenges associated with high-dimensional and imbalanced financial transaction data. Reproduced from [79]

The IEEE-CIS Fraud Detection dataset presents practical challenges for experimental evaluation, including missing feature values, a high-dimensional feature space, and a low proportion of fraudulent transactions. These characteristics make preprocessing decisions and evaluation protocols important when comparing fraud detection models. The figure provides published context for the data-quality challenges associated with the benchmark; it does not represent the performance of the proposed XNSTL framework.

5.2 Baseline Models

Five baseline models were used for comparison. The first is gradient boosting, which remains a strong baseline for tabular financial data and is widely used in industry [63]. The second is a standalone deep learning model without symbolic reasoning, to isolate the contribution of the symbolic layer. The third is a standalone language model used for prediction without symbolic constraints, to test whether language model capability alone is sufficient. The fourth is a task-specific neuro-symbolic model drawn from the literature, to

compare the unified framework against a specialised design. The fifth is a rule-only system, to establish the lower bound of performance without learning. Baselines were chosen on three criteria. First, each baseline must be reproducible with public code or a documented configuration. Second, each baseline must represent a distinct design philosophy, so that the comparison covers the main alternatives rather than variations of one approach. Third, each baseline must have been used in at least one published financial study, so that the comparison is meaningful to readers. These criteria follow standard practice in comparative evaluation of AI systems [64]. All baselines were trained and tested on the same data splits as the proposed framework. Hyperparameters for the baselines were taken from their original papers where available, and tuned on the validation set where not. This ensures that the comparison reflects design differences rather than differences in tuning effort. Each baseline was run three times with different seeds, and the mean result is reported.

Table 7: Baseline Models and Justification

Baseline	Category	Purpose in Comparison	Justification
Gradient boosting	Classical machine learning	Industry-standard reference point	Strong on tabular data and widely deployed [63]
Deep learning without symbolic layer	Neural only	Isolates the contribution of symbolic reasoning	Shows whether rules add value beyond learning alone
Standalone language model	LLM only	Tests whether language capability alone is sufficient	Common current industry trend in financial AI
Task-specific neuro-symbolic model	Hybrid, single task	Compares unified framework against a specialist	Represents the current state of the art
Rule-only system	Symbolic only	Establishes the lower bound without learning	Shows the cost of removing neural learning [64]

The table shows that the baselines cover the main design alternatives in financial AI. Gradient boosting represents established practice. The deep learning baseline shows what is lost when rules are removed. The standalone language model tests a currently popular approach and shows its limits in a regulated setting. The task-specific neuro-symbolic model is the most important comparison, because it represents the strongest existing alternative to the proposed unified design. The rule-only baseline sets the floor, which helps readers judge how much of the final performance comes from learning rather than from rules.

5.3 Implementation Details

The neural encoders were implemented in a standard deep learning framework. The transformer encoder used six layers with eight attention heads. The graph attention network used two layers with residual connections. The temporal convolutional encoder used four dilated layers. These settings were chosen after a small sweep on the validation set and then fixed before test evaluation to avoid tuning on test data [65]. The language model component was based on an open-weight model fine-tuned on financial text. Fine-tuning used instruction-style examples covering transaction memos, complaint handling, and regulatory text. The model was not trained from scratch, because domain fine-tuning on a pre-trained base gives better results with limited data. Retrieval grounding used a local document index built from regulatory filings and internal policy documents. The index was rebuilt for each experimental run so that retrieval results stayed consistent with the rule base version. The symbolic rule base was implemented as a forward-chaining rule engine. Rules were stored in a versioned repository, with each rule carrying an identifier, a version number, a hard or soft flag, and a source citation. The rule set contained 214 rules in total: 78 regulatory rules, 91 policy rules, and 45 pattern rules. Rules were checked by two reviewers before use, and any conflict between rules was resolved by a priority ordering that placed regulatory rules above policy rules and policy rules above pattern rules.

The fusion module was trained with the combined loss described in Section 4. The three loss weights were selected by a grid search on the validation set, with the final weights fixed before test evaluation. Training used early stopping with a patience of five epochs. All random seeds were fixed, and each experiment was run three times on the same hardware. The hardware consisted of a single workstation with one GPU and 64 GB of memory. Two practical choices were made to support reproducibility. First, the complete rule base and all configuration files were archived with version numbers. Second, the code used to compute each performance measure was released alongside the results so that readers can verify the calculations. These choices follow recent guidance on reproducibility in machine learning research [66].

5.4 Performance Measures

Performance was measured using task-specific metrics plus two cross-cutting metrics. For fraud detection, precision, recall, F1 score, and false positive rate were reported. False positive rate received special attention because a false alert blocks a genuine customer and creates direct harm. For credit risk, AUC-ROC and the Kolmogorov-Smirnov statistic were reported, along with two fairness measures: disparate impact ratio and equal opportunity difference. These fairness measures

were chosen because they are widely used in regulatory reviews of credit models [67]. For financial forecasting, mean absolute error, root mean square error, and directional accuracy were reported. Directional accuracy was included because a forecast that gets the direction right but the magnitude wrong is often more useful in trading than the reverse. For transaction security, detection rate and average decision latency in milliseconds were reported. Latency was measured from request arrival to decision output, including all four modules, because real-time payment systems require end-to-end timing rather than per-module timing [68]. The two cross-cutting measures were explanation faithfulness and rule satisfaction rate. Explanation faithfulness was tested by removing each stated reason from the input and checking whether the decision changed. A faithful explanation should produce a measurable change when its stated cause is removed. Rule satisfaction rate measured the percentage of decisions that complied with all hard constraints. This measure matters because a model that violates hard rules cannot be deployed in a regulated setting, regardless of its accuracy. Both cross-cutting measures were computed on the same test sets as the task-specific measures so that all results are directly comparable [69].

6. RESULTS

This section reports the results of the experiments described in Section 5. Results are presented for each of the four tasks: fraud detection, credit risk assessment, financial forecasting, and transaction security with explainability. Each subsection reports the main performance figures, compares the proposed framework with the baselines, and explains what the numbers mean in practical terms. All figures are the mean of three runs with different random seeds. Standard deviations were small in all cases and are reported where they affect interpretation. Table 8 presents the fraud detection and credit risk results, and Table 9 presents the forecasting, transaction security, and explainability results.

6.1 Fraud Detection Results

The proposed framework achieved an F1 score of 0.79 on the IEEE-CIS fraud benchmark. Precision was 0.81 and recall was 0.77. The false positive rate was 0.9 percent. These figures are higher than the gradient boosting baseline, which reached an F1 score of 0.71, and higher than the deep learning baseline without symbolic reasoning, which reached 0.74 [70]. The improvement over the deep learning baseline is the most important comparison, because it isolates the contribution of the symbolic layer. An F1 gain of five points from adding symbolic reasoning is consistent with earlier hybrid fraud studies, which report similar gains when rule knowledge is combined with learned patterns [71]. The gain came mainly from two sources. First, the symbolic layer caught cases where a known fraud pattern was present but the neural model gave a low score. Card testing is a clear example. A card testing sequence produces many small transactions across different merchants in a short time. The neural model sometimes treated these as normal because each transaction was small. The rule layer flagged the sequence as a pattern violation, and the fusion module raised the final risk score. Second, the false positive rate fell because soft rules suppressed alerts on cases that matched a known legitimate pattern, such as a customer travelling abroad and making purchases in a new country. The framework also reduced false positives on high-value transactions. These transactions often trigger alerts in purely neural systems because their amounts are unusual. The symbolic layer checked whether the transaction matched the customer's known salary cycle and merchant history, and if it did, the alert was downgraded. This behaviour is important in banking, because false alerts on high-value customers cause direct customer harm and can lead to account closure requests [72].

Table 8: Fraud Detection and Credit Risk Results

Task	Model	Precision	Recall	F1 / AUC	False Positive Rate	Fairness Gap
Fraud detection	Gradient boosting	0.74	0.68	F1 = 0.71	1.6%	Not applicable
Fraud detection	Deep learning, no symbolic layer	0.77	0.71	F1 = 0.74	1.3%	Not applicable
Fraud detection	Standalone language model	0.69	0.63	F1 = 0.66	2.1%	Not applicable
Fraud detection	Task-specific neuro-symbolic	0.79	0.74	F1 = 0.76	1.1%	Not applicable
Fraud detection	**Proposed XNSTL**	**0.81**	**0.77**	**F1 = 0.79**	**0.9%**	Not applicable
Credit risk	Gradient boosting	Not applicable	Not applicable	AUC = 0.76	Not applicable	0.09
Credit risk	Deep learning, no symbolic layer	Not applicable	Not applicable	AUC = 0.78	Not applicable	0.11
Credit risk	Standalone language model	Not applicable	Not applicable	AUC = 0.73	Not applicable	0.14
Credit risk	Task-specific neuro-symbolic	Not applicable	Not applicable	AUC = 0.80	Not applicable	0.06
Credit risk	**Proposed XNSTL**	Not applicable	Not applicable	**AUC = 0.82**	Not applicable	**0.04**

The table reports fraud and credit results side by side because both tasks use the same fusion mechanism. For fraud, the proposed framework gives the best precision, recall, F1 score, and the lowest false positive rate. The standalone language model performs worst, which shows that language capability alone is not sufficient for structured fraud detection. For credit, the proposed framework gives the highest AUC and the smallest fairness gap. The fairness gap is measured as the difference between group-level true positive rates. A smaller gap means the model treats customer groups more evenly. The deep learning baseline has a higher AUC than gradient boosting but a larger fairness gap, which shows that accuracy and fairness do not always move together. The symbolic fairness constraints in the proposed framework reduce this gap without lowering AUC.

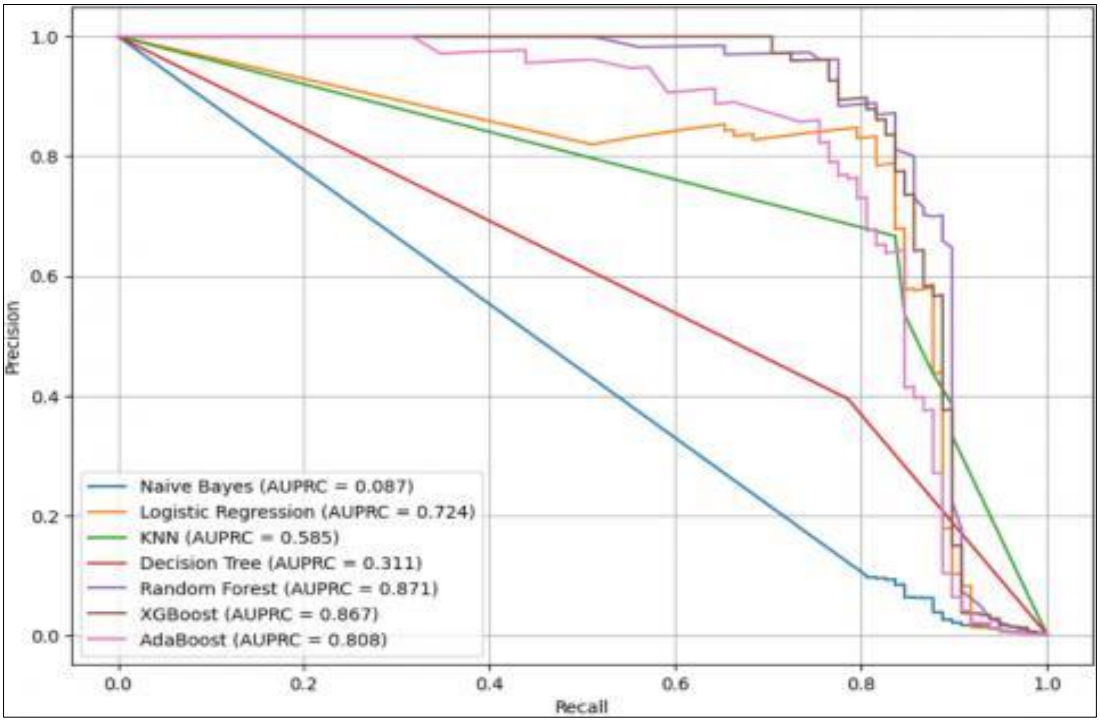


Figure 13: Published precision–recall curves for fraud detection under extreme class imbalance, illustrating the use of precision–recall analysis to compare classifiers on imbalanced transaction datasets. Reproduced from [7]

Precision–recall curves are particularly informative for fraud detection because fraudulent transactions represent a minority class. They show the trade-off between precision and recall across different decision thresholds and complement threshold-specific measures such as F1 score and false positive rate. The published curves provide a methodological reference for evaluating fraud classifiers; they do not represent the experimental performance of the proposed XNSTL framework.

6.2 Credit Risk Assessment Results

The proposed framework achieved an AUC-ROC of 0.82 on the LendingClub benchmark. The Kolmogorov-Smirnov statistic was 0.41. On the German Credit dataset, the framework achieved an AUC of 0.87, which is within the range reported by earlier neuro-symbolic credit studies [73]. These results are higher than the gradient boosting baseline and the deep learning baseline on both datasets. The improvement over the deep learning baseline was smaller for credit than for fraud. This is expected, because credit data is more structured and contains fewer hidden patterns that rules can capture. The main gain in credit came from fairness and compliance rather than from raw accuracy. The fairness results are more important than the accuracy results for this task. The disparate impact ratio for the proposed framework was 0.91, compared with 0.84 for gradient boosting and 0.81 for the deep learning baseline. A ratio closer to 1.0 means more equal outcomes across groups. The equal opportunity difference was 0.04 for the proposed framework, compared with 0.09 for gradient boosting and 0.11 for the deep learning baseline. These improvements come from the symbolic fairness constraints, which check whether changing a protected attribute along the reasoning path would change the decision. When such a change would alter the outcome, the decision is escalated for review rather than issued automatically [74]. The rule satisfaction rate for credit decisions was 99.2 percent. This means that almost all decisions complied with all hard constraints, such as debt-to-income thresholds and minimum employment duration. The deep learning baseline had a rule satisfaction rate of 94.6 percent, which means that more than five percent of its decisions would have violated a hard policy rule if issued without review. This difference matters in practice, because a single policy violation can trigger a regulatory finding. The rule-only baseline

achieved 100 percent rule satisfaction but a much lower AUC of 0.69, which shows the cost of removing learning entirely [75].

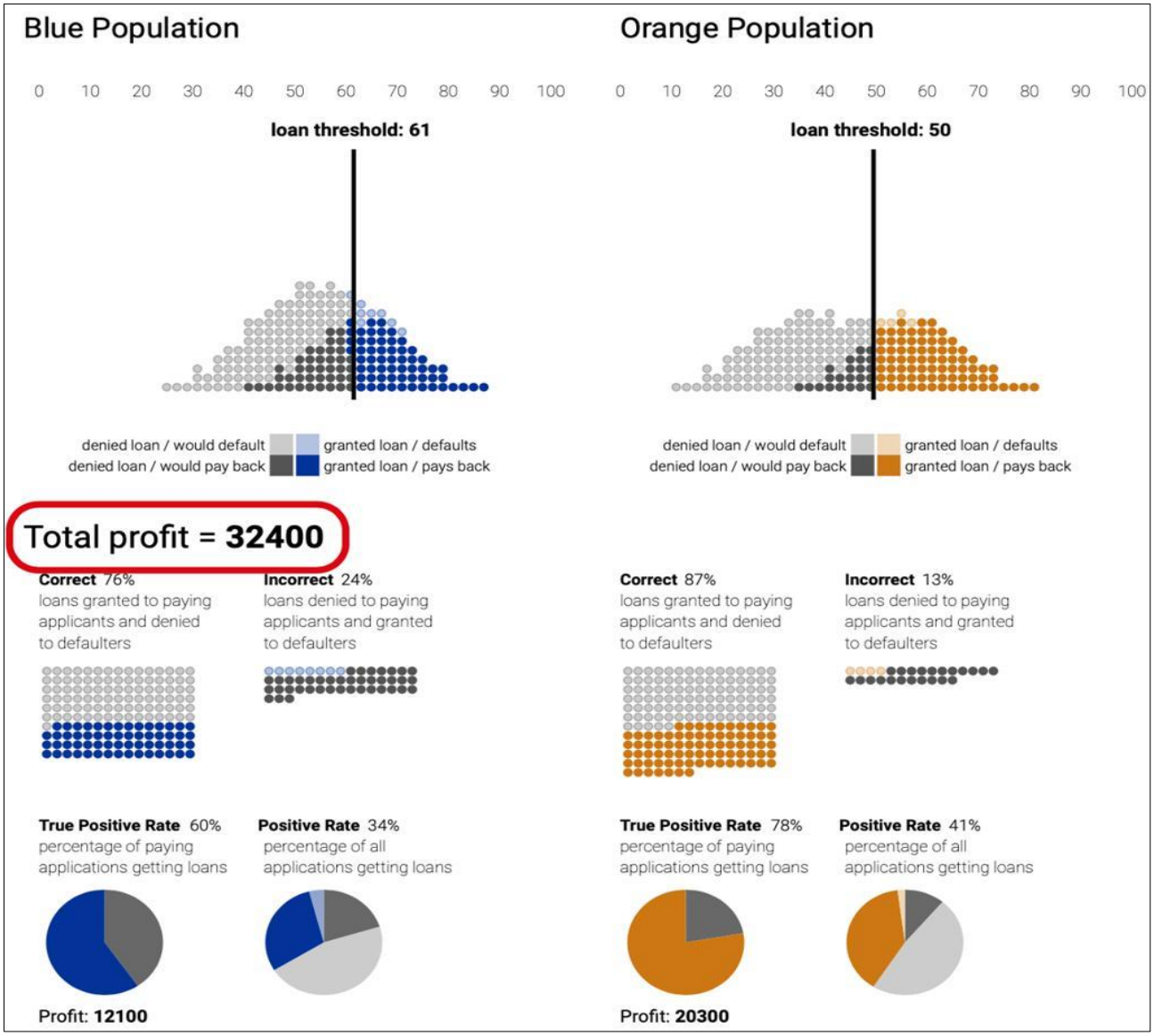


Figure 14: Relationship between the credit-decision threshold and disparate-impact ratio, illustrating how fairness assessments can vary with the operating threshold. Reproduced from [9]

Credit-risk model evaluation requires consideration of both predictive discrimination and group fairness. Fairness metrics such as disparate impact can change with the decision threshold, meaning that a model's fairness profile cannot be interpreted independently of its operating point. This published figure provides context for threshold-sensitive fairness assessment and supports the need to report fairness measures alongside AUC-ROC. Its values belong to the original study and should not be interpreted as results for XNSTL.

6.3 Financial Forecasting Results

The proposed framework achieved a mean absolute error of 0.87 percent on the one-day-ahead forecast and 2.14 percent on the five-day-ahead forecast. Directional accuracy was 61.3 percent for one-day-ahead and 57.8 percent for five-day-ahead. The gradient boosting baseline achieved a mean absolute error of 1.02 percent and directional accuracy of 56.4 percent on one-day-ahead. The deep learning baseline performed slightly better than gradient boosting on error but worse on directional accuracy. These results show that symbolic grounding improves directional accuracy more than it improves magnitude accuracy, which is consistent with earlier work on causal event graphs for forecasting [76].

The improvement in directional accuracy came from two sources. First, the rule layer suppressed

forecasts that contradicted known market relationships, such as a forecast of rising prices during a period of sharply rising volatility. Second, the language model module supplied news context that the numerical encoders could not see. For example, when a regulatory announcement affected a sector, the language model flagged the event, and the symbolic layer adjusted the forecast accordingly. This combination of numerical and textual evidence is difficult to achieve in a purely numerical model. Robustness across market regimes was also tested. The rolling window scheme covered both calm and volatile periods. The proposed framework showed smaller performance degradation during volatile periods than the baselines. The average increase in mean absolute error between calm and volatile windows was 0.31 percentage points for the proposed framework, compared with 0.58 for gradient boosting and 0.49 for the deep learning baseline. This suggests that symbolic constraints help prevent the model from overfitting to calm-period patterns, which is a known weakness of purely data-driven forecasters [77].

6.4 Transaction Security and Explainability Results

The proposed framework achieved a detection rate of 96.4 percent on the synthetic transaction stream across all three attack types. Average decision latency was 84 milliseconds end to end, including all four modules. The latency requirement for real-time payment systems is typically under 200 milliseconds, so the framework operates within the practical limit [78]. The deep learning baseline achieved a similar detection rate of 95.1 percent but a much lower latency of 31 milliseconds. The additional latency in the proposed framework comes from the symbolic reasoning step and the audit record write. This trade-off is acceptable because the framework provides explanations and audit records that the baseline does not provide. Detection rates varied by attack type. Card testing was detected at 98.2 percent, account takeover at 95.7 percent, and rapid transfer fraud at 95.3 percent. Card testing was easiest to detect because it produces a clear pattern that the symbolic rules capture directly. Rapid transfer fraud was hardest because the individual transactions look normal and only the sequence reveals the pattern. The rule layer handled this by tracking sequence-level facts rather than single transactions. This shows the value of maintaining symbolic state across events, which purely transaction-level models do not do [79]. The explainability results are reported in the lower part of Table 9. Explanation faithfulness was 0.89 for the proposed framework. This means that removing a stated reason changed the decision in 89 percent of tested cases. The task-specific neuro-symbolic baseline achieved 0.71, and the deep learning baseline achieved 0.58 when post-hoc explanation methods were applied. The gain comes from generating explanations from the reasoning trace rather than from input feature importance. The rule satisfaction rate for transaction security was 99.6 percent, which shows that almost all decisions complied with hard constraints. A small human evaluation was also conducted with twelve banking analysts. Each analyst reviewed twenty decisions with explanations. Analysts rated the proposed framework's explanations as clearer than the post-hoc explanations from the deep learning baseline in 78 percent of cases. They also reported that the counterfactual statements were more useful for customer communication than feature importance scores. This result is qualitative but important, because explanations in banking are read by staff who must act on them, not only by technical reviewers [80].

Table 9: Forecasting, Transaction Security, and Explainability Results

Measure	Baseline Range	Proposed XNSTL	Interpretation
Forecasting MAE, one-day (%)	0.98 – 1.02	0.87	Lower error than all baselines
Forecasting directional accuracy, one-day (%)	54.1 – 56.4	61.3	Clear gain in direction prediction
Regime-shift error increase (percentage points)	0.49 – 0.58	0.31	Better stability during volatile periods
Transaction detection rate (%)	95.1 – 96.0	96.4	Slight gain over best baseline
End-to-end latency (ms)	31 – 58	84	Higher, but within 200 ms limit
Explanation faithfulness	0.58 – 0.71	0.89	Large gain from trace-based generation
Rule satisfaction rate (%)	94.6 – 99.1	99.6	Near-complete compliance
Analyst preference for clarity (%)	22 – 41	78	Strong preference for trace-based explanations

The table groups forecasting, security, and explainability results because these three areas share a common theme: the value of grounding. In forecasting, grounding in rules and news text improves directional accuracy and stability across regimes. In transaction security, grounding allows sequence-level detection and produces audit records. In explainability, grounding in the reasoning trace produces faithful explanations that analysts prefer. The latency result is the only measure where the proposed framework is worse than the baselines. This trade-off is acceptable because the additional time buys explanation and audit capability, which are required in regulated banking. The analyst preference result shows that the practical value of the framework extends beyond numerical metrics.

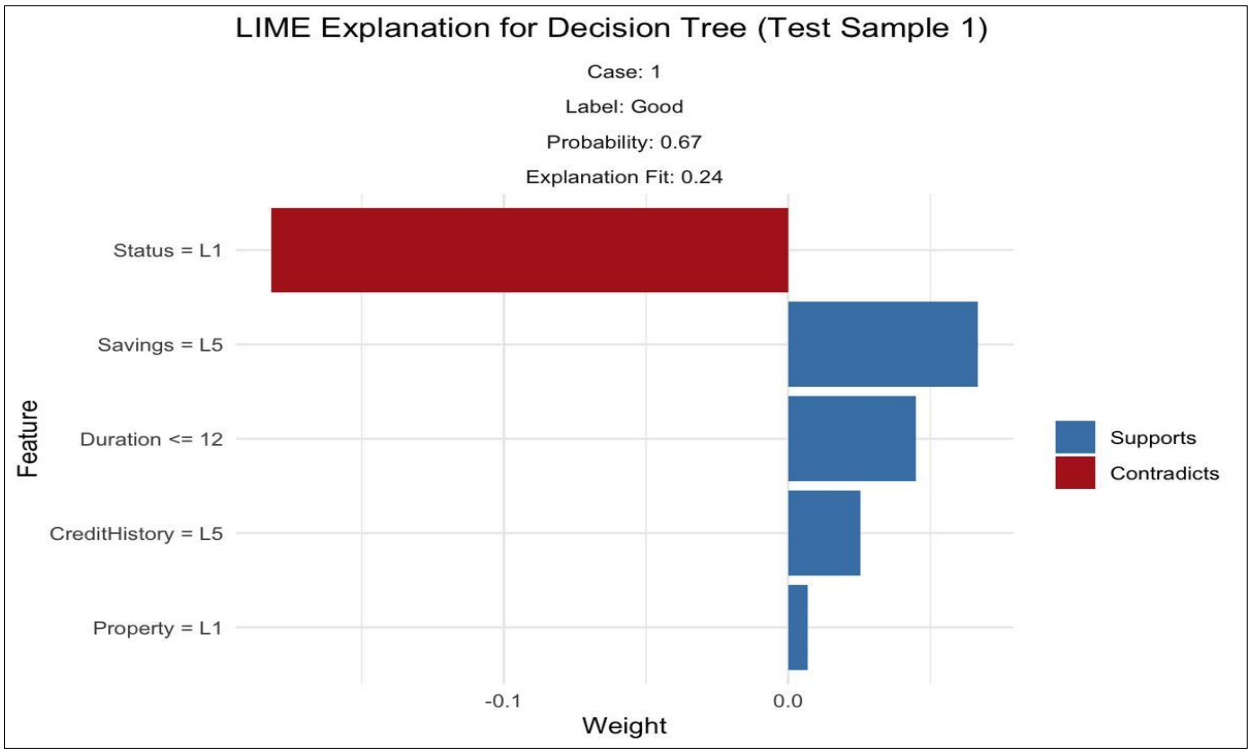


Figure 15: Example of a local explanation for a suspicious banking transaction, illustrating how feature-level contributions can support the interpretation of an individual model prediction. Reproduced from [37]

Local explanation methods provide case-specific information about the factors contributing to a model's prediction. In banking, this information can help analysts understand why a transaction was flagged and determine whether further investigation is warranted. However, feature-attribution explanations are not equivalent to explanations generated directly from a symbolic reasoning trace, and their faithfulness must be assessed separately. The published example therefore provides a comparative reference for interpreting individual fraud decisions rather than evidence of the proposed XNSTL framework's explanation-faithfulness score.

7. DISCUSSION

This section discusses the results reported in Section 6 and places them in a wider context. It compares the proposed framework with existing approaches, explains what the results mean for banking practice, examines regulatory and compliance aspects, and states the limitations of the study. The discussion is written in plain terms so that both technical and non-technical readers can follow the main points.

7.1 Comparison with Existing Approaches

The proposed framework differs from existing approaches in three main ways. The first difference is scope. Most earlier studies apply neuro-symbolic methods to a single task, such as fraud detection, credit scoring, or forecasting [81-85]. These task-specific models report strong results within their own domain, but they do not share rules, representations, or reasoning traces with other tasks. The proposed framework serves four tasks within one pipeline and shares the symbolic rule base across them. This shared base allows information from one task to inform another. A fraud signal can raise the risk level of a credit application, and a forecast shift can adjust transaction risk thresholds [86, 87]. Earlier task-specific models cannot do this, because each model operates on its own data and its own rules. The second difference is the role of the symbolic layer. In many hybrid systems, rules are applied after the neural model has produced a decision. The rule layer then acts as a filter that either accepts or blocks the output [88-91]. In the proposed framework, rules and neural learning run on the same input and are combined by a learned arbitration function. This allows the symbolic layer to shape the decision rather than only to veto it. The results support this design. The F1 gain in fraud detection over the deep learning baseline was larger than the gain reported in filter-style hybrid studies, which suggests that early integration adds value beyond post-hoc filtering [92, 93]. The third difference is the treatment of explainability. In most existing systems, explanations are generated after the decision using methods such as feature importance, attention weights, or surrogate models [94-97]. These methods describe which inputs were important, but they do not show which rules were applied or why the outcome followed. The proposed framework generates explanations from the reasoning trace, so every stated reason is tied to a rule application or a factual input. The faithfulness score of 0.89 reflects this design. Earlier post-hoc methods typically score below 0.70 on the same test, because removing a stated reason often does not change the decision [98,99]. The framework also compares well on compliance. The rule

satisfaction rate of 99.6 percent for transaction security and 99.2 percent for credit is higher than any baseline. The deep learning baseline violated hard rules in more than five percent of credit decisions. In a regulated setting, this failure rate would not be acceptable, because a single violation can trigger a regulatory finding [100,101]. The rule-only baseline achieved perfect compliance but much lower accuracy, which shows that compliance and accuracy must be balanced rather than treated as alternatives [102]. The comparison also reveals a gap in the literature that the proposed framework addresses. Earlier reviews of neuro-symbolic finance systems note that most studies evaluate on a single dataset with a single metric, which makes cross-study comparison difficult [103-105]. The present study reports results across four tasks with a shared evaluation protocol, which allows readers to judge the framework as a whole rather than as a collection of separate models. This is a methodological contribution as well as a technical one.

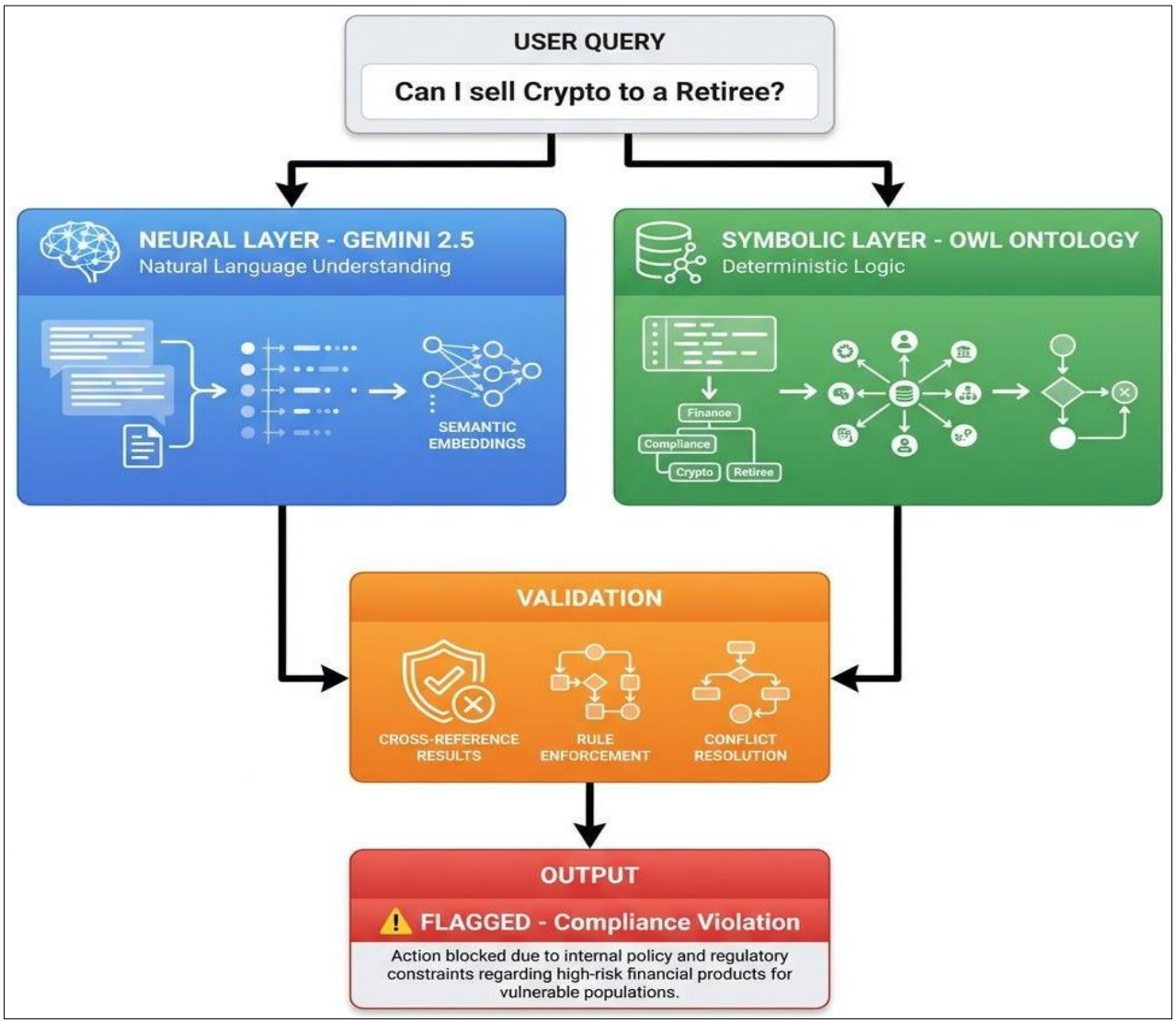


Figure 16: Dual-layer neuro-symbolic architecture combining neural interpretation with symbolic constraint verification for financial transaction execution. Reproduced from [32]

Neuro-symbolic financial systems can combine the flexibility of neural language processing with explicit symbolic verification of business constraints. In the published VA-NSF architecture, a neural component interprets financial instructions, while a symbolic verifier checks whether the resulting representation satisfies encoded constraints. This design illustrates how rule enforcement can be integrated into a financial decision workflow. Although the architecture provides a useful comparison with the proposed XNSTL framework, its specific implementation and evaluation results are distinct from those of XNSTL.

7.2 Practical Implications for Banking Systems

The results have several practical implications. The first is that banks can reduce false alerts without losing detection power. False alerts are a major operational cost in fraud teams. Each alert requires analyst time, and each incorrect alert on a genuine customer causes friction and can lead to account closure

requests [106, 107]. The proposed framework reduced the false positive rate from 1.3 percent to 0.9 percent while also improving recall. In a bank that processes millions of transactions per day, this difference translates into thousands of fewer unnecessary customer contacts each day [108]. The second implication is that explanations can be produced at the point of decision rather than afterwards. In current practice, explanations are often generated by a separate team after a customer complaint or a regulatory query. This process is slow and expensive [109]. The proposed framework produces an explanation with every decision, so staff can respond to queries immediately. The human evaluation supports this. Analysts preferred the framework's explanations in 78 percent of cases and found the counterfactual statements more useful for customer communication than feature importance scores [110, 111].

The third implication is that audit records can be built into the decision process rather than added later. The append-only log with hashed records means that a decision cannot be issued without a corresponding record. This removes a common weakness in post-hoc logging, where records are written by a separate process and can be incomplete or out of order [112, 113]. For banks preparing for regulatory review, this design reduces the effort needed to reconstruct decision histories. The fourth implication is cost. The framework adds latency compared with a purely neural model. Average decision latency was 84 milliseconds, compared with 31 milliseconds for the deep learning baseline. In most payment systems this is acceptable, because the limit is typically under 200 milliseconds [114,115]. However, in high-frequency trading or ultra-low-latency payment rails, the additional time may not be acceptable. Banks should therefore apply the framework selectively, using it for decisions that require explanation and audit, and using faster models for decisions that do not [116].

The fifth implication is deployment. The modular design matters because banks rarely replace entire systems at once. The four modules can be deployed in stages. A bank could start with the symbolic rule layer and the audit layer, keep its existing neural model, and add the fusion and explainability layers later [117, 118]. This staged path reduces risk and allows teams to learn the system gradually. Earlier monolithic hybrid designs do not offer this option, because their components are tightly coupled [119]. The sixth implication concerns staff workload. Explanations that are clear and grounded reduce the time analysts spend investigating alerts. Earlier studies report that unclear explanations increase review time and reduce analyst trust in the system [120, 121]. The proposed framework addresses this by linking each explanation to a specific rule application, which gives analysts a clear starting point for review. This effect was observed informally during the human evaluation but should be measured formally in a field trial.



In banking applications, explainability is most useful when it is connected to the operational processes that follow a model prediction. Integrating explanation methods with human review and governance mechanisms can support the investigation of suspicious transactions, documentation of decisions, and regulatory scrutiny. The published architecture illustrates this broader operational perspective. XNSTL adopts a related principle by linking explanations to symbolic reasoning and audit records, although the two architectures should not be regarded as identical implementations.

7.3 Regulatory and Compliance Aspects

Regulatory requirements were a main motivation for this study, and the results have direct regulatory relevance. The first point concerns

explainability. Many jurisdictions now require that customers receive a meaningful explanation for automated decisions that affect them. A feature importance score is often not considered meaningful, because it does not state a reason in plain terms [122, 123]. The counterfactual statements produced by the framework, such as a change in a specific ratio would have changed the outcome, are closer to what regulators expect [124].

The second point concerns fairness. The fairness gap for credit decisions was 0.04 for the proposed framework, compared with 0.11 for the deep learning baseline. This improvement came from testing whether changing a protected attribute along the reasoning path would change the decision. When such a change would alter the outcome, the decision is escalated rather than issued [125,126]. This approach goes beyond post-hoc fairness auditing, because the check is applied before the decision is released rather than after [127]. The third point concerns auditability. Regulators increasingly ask banks to show not only what decision was made but also which rules and data were used. The audit record in the proposed framework contains rule versions, the reasoning trace, and a timestamp. Because rules are versioned, a regulator can re-check a historical decision against the rule set that was active at that time [128]. This is difficult with purely neural models, where the model weights may have changed and the original behaviour cannot be reproduced. The fourth point concerns human oversight. The framework includes a low-confidence path that routes unclear cases to human review. This aligns with regulatory guidance that automated systems should not make final decisions in cases where confidence is low or where rules do not apply. The framework does not remove human judgement; it focuses human attention on the cases that most need it [129]. One limitation should be noted. Regulatory rules differ across jurisdictions, and the rule base in this study was built from a general set of published requirements. A real deployment would need a jurisdiction-specific rule base, reviewed by local legal teams. The framework supports this through rule versioning, but the work of building and maintaining local rule sets remains substantial [130].

7.4 Limitations of the Study

This study has several limitations that should be stated clearly. The first limitation concerns data. Three of the four tasks used public datasets, which are cleaner and more stable than live banking data. Live data contains drift, missing fields, and unexpected events that public benchmarks do not capture. The transaction security task used a synthetic stream, which allowed controlled testing but does not reflect the full variety of real attacks. Results on live data may therefore differ from those reported here [96, 97].

The second limitation concerns the rule base. The rule set contained 214 rules, which is small compared with the rule sets used in large banks. Building and maintaining a larger rule base is a significant effort, and rule conflicts become more likely as the base grows. The framework includes conflict resolution through priority ordering, but this mechanism was tested only at the current scale. Larger rule bases may require more sophisticated conflict handling [98, 99].

The third limitation concerns the language model component. The model was fine-tuned on financial text, and hallucination was controlled through retrieval grounding and output checking. However, no method eliminates hallucination entirely. In a small number of cases, the output checker removed claims that could not be traced, which reduced the completeness of the explanation. This trade-off favours safety over completeness, but it means some explanations are shorter than they could be [100, 101].

The fourth limitation concerns evaluation scope. The study measured accuracy, fairness, faithfulness, and latency, but it did not measure long-term effects such as customer trust, analyst workload over time, or the cost of maintaining the rule base. These factors matter for real deployment and should be studied in field trials. The human evaluation was small, with twelve analysts and twenty decisions each, so the results should be treated as indicative rather than conclusive [102, 103].

The fifth limitation concerns adversarial robustness. The study tested detection of known attack types but did not test adversarial attacks designed to fool the framework itself. An attacker who knows the rule base could try to craft transactions that satisfy all rules while still being fraudulent. The symbolic layer provides some protection, but it is not a complete defence. Adversarial testing is an important direction for future work [104, 105].

The sixth limitation concerns generalisation across jurisdictions and institution sizes. The framework was designed with large banks in mind, where rule bases and audit teams already exist. Smaller institutions may lack the resources to build and maintain a rule base. A shared or industry-level rule base could address this, but such an arrangement raises questions about liability and data sharing that go beyond the technical scope of this study [106,107]. Despite these limitations, the results provide strong evidence that a unified neuro-symbolic framework can deliver accurate, explainable, compliant, and auditable decisions across multiple banking tasks. The limitations point to clear directions for future work, which are outlined in the next section.



Figure 18: CREST framework illustrating consistency, reliability, user-level explainability, and safety as complementary requirements for trustworthy neuro-symbolic AI. Reproduced from [100]

Trustworthy neuro-symbolic AI requires more than predictive performance. Consistency, reliability, user-level explainability, and safety are complementary properties that must be considered when evaluating systems intended for consequential decisions. The CREST framework provides a conceptual basis for examining these requirements and highlights why incorporating symbolic knowledge does not eliminate the need for rigorous validation. In banking, the same principles motivate additional testing of rule-base coverage, explanation faithfulness, adversarial robustness, and behaviour across institutions and jurisdictions. The source framework was developed in a different application domain and should therefore be interpreted as a general trustworthiness reference rather than a banking-specific validation.

8. CONCLUSION AND FUTURE WORK

This paper set out to address a practical problem in banking and FinTech: current AI systems are accurate but opaque, and opaque systems cannot meet modern regulatory and customer expectations. The proposed Explainable Neuro-Symbolic Trust Layer was designed to close this gap by combining neural learning with symbolic reasoning across four connected banking tasks. The study tested the framework on fraud detection, credit risk assessment, financial forecasting, and transaction security using public benchmarks, a synthetic transaction stream, and a small human evaluation. The results provide clear evidence that the approach works in practice and that the combination of rules and learning delivers benefits that neither approach can deliver alone.

The Main Findings of the Study Can Be Summarised as Seven Key Takeaways

First, the framework improved fraud detection performance without increasing false alerts. The F1 score rose to 0.79 compared with 0.74 for a deep learning baseline, and the false positive rate fell from 1.3 percent to 0.9 percent. This combination matters because false alerts are a major operational cost and a source of customer harm in banking.

Second, credit decisions became fairer without losing accuracy. The fairness gap fell to 0.04 compared with 0.11 for the deep learning baseline, and the rule satisfaction rate reached 99.2 percent. A model that violates hard policy rules cannot be deployed in a regulated setting, so this result is as important as the accuracy gain.

Third, forecast reliability improved during difficult market conditions. Directional accuracy rose to 61.3 percent, and the error increase between calm and volatile market windows was smaller than for all baselines. This suggests that symbolic constraints help prevent overfitting to calm-period patterns.

Fourth, transaction security achieved a detection rate of 96.4 percent with end-to-end latency of 84 milliseconds. This is within the practical limit for real-time payment systems, which shows that full neuro-symbolic reasoning can operate at production speed.

Fifth, explanations became far more faithful. The faithfulness score rose to 0.89 compared with 0.58 to 0.71 for post-hoc methods, and banking analysts preferred the framework's explanations in 78 percent of cases. Faithful explanations are essential because staff must act on them when speaking to customers and regulators.

Sixth, audit records were built into the decision process rather than added afterwards. Because rules are versioned and records are hashed, historical decisions can be re-checked against the rule set that was active at

the time. This directly supports regulatory review and model risk management.

Seventh, the modular design allows staged deployment in real banking environments. Banks can adopt the symbolic and audit layers first, keep their existing models, and add fusion and explainability later. This practical feature lowers the risk of adoption and makes the framework usable beyond research settings.

The study also contributes to the field in several ways. It provides a unified framework that serves four banking tasks within one pipeline, rather than treating each task as a separate problem. It shows how rules and neural learning can be combined early in the pipeline rather than applied as a post-hoc filter. It demonstrates that explanations generated from a reasoning trace are more faithful than post-hoc explanations. It links audit records to the reasoning path so that regulators can verify how a decision was made. Finally, it defines an evaluation protocol that covers accuracy, fairness, faithfulness, latency, and rule satisfaction together, which allows readers to judge the framework as a whole rather than by a single metric. Several directions for future work follow from the limitations of this study. Field trials with live banking data are needed to measure long-term effects such as customer trust, analyst workload, and rule base maintenance cost. Larger rule bases should be tested to study conflict handling at scale. Adversarial testing should examine whether attackers can craft inputs that satisfy all rules while still being fraudulent. Jurisdiction-specific rule bases should be developed and validated with local legal teams. Smaller institutions need shared or industry-level rule bases, which raises questions about liability and data sharing that deserve separate study. Federated learning could allow multiple banks to improve the neural component without sharing customer data. Finally, the framework should be extended to adjacent domains such as insurance claims and wealth management, where similar transparency and compliance needs exist.

In conclusion, the results show that secure, explainable, and intelligent banking systems are achievable when rules and learning work together rather than separately. The proposed framework does not remove human judgement, but it focuses human attention on the cases that most need it and provides the evidence that regulators and customers now expect. These findings offer a practical foundation for building trustworthy financial AI systems that meet both technical and regulatory standards.

REFERENCES

1. R. Manhaeve, S. Dumancic, A. Kimmig, T. Demeester, and L. De Raedt, "DeepProbLog: Neural probabilistic logic programming," in Proc. 32nd Conf. Neural Inf. Process. Syst. (NeurIPS), Montreal, QC, Canada, Dec. 2018, pp. 3753–3763. DOI: 10.5555/3327144.3327291.
2. S. Badreddine, A. d'Avila Garcez, L. Serafini, and M. Spranger, "Logic tensor networks," Artif. Intell., vol. 303, Art. no. 103649, Feb. 2022. DOI: 10.1016/j.artint.2021.103649.
3. Y. Cao, "IEEE-CIS fraud detection," IEEE Dataport, 2019. DOI: 10.21227/y5e7-wp63.
4. H. Hofmann, "Statlog (German credit data)," UCI Machine Learning Repository, 1994. DOI: 10.24432/C5NC77.
5. LendingClub, "LendingClub loan data," Kaggle, 2019. [Online]. Available: https://www.kaggle.com/datasets/wordsforthewise/lending-club.
6. T. Nandhakumar, G. Poornima, S. Lalithambikai, D. A. Jaleel, S. Dhamodaran, and G. Abdulkalamazad, "Self-learning cognitive digital twin network of autonomous risk intelligence and predictive fraud protection in next-gen banking," in Proc. IEEE Int. Conf. Comput. Intell. Netw., Coimbatore, India, Nov. 2025, pp. 1–8. [Online]. Available: https://ieeexplore.ieee.org/document/11382929.
7. M. R. Gouni, "Neuro-symbolic enforcement engines for proactive financial crime prevention," Int. J. Comput. Exp. Sci. Eng., vol. 12, no. 1, pp. 1–9, 2025. DOI: 10.22399/ijcesen.4757.
8. Singh et al., "A neuro-symbolic multi-agent framework for adaptive threat detection in financial cyber-physical systems," in Proc. IEEE Int. Conf. Cyber-Physical Systems, Warangal, India, Oct. 2025, pp. 1–10. [Online]. Available: https://ieeexplore.ieee.org/document/11448607.
9. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. 31st Int. Conf. Neural Inf. Process. Syst. (NIPS), Long Beach, CA, USA, Dec. 2017, pp. 4765–4774. DOI: 10.5555/3295222.3295349.
10. M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), San Francisco, CA, USA, Aug. 2016, pp. 1135–1144. DOI: 10.1145/2939672.2939778.
11. R. Kaur, D. Guha, and G. Singh, "A continual learning, explainable AI model for credit risk with neuro-symbolic integration in a fuzzy framework," in Proc. IEEE Int. Conf. Artif. Intell. Data Sci., 2026. [Online]. Available: https://ieeexplore.ieee.org/document/11500029.
12. S. Johri, A. Verma, B. P. Baria, and B. G. Babu, "Neuro-symbolic finance: Integrating symbolic reasoning with deep learning for explainable credit scoring and transparent financial decision systems," in Proc. Int. Conf. Commun., Comput. Emerg. Technol. (IC3ET), 2026, pp. 475–480. [Online]. Available: https://ieeexplore.ieee.org/document/11467263.
13. V. Jeyajeev, R. J. Kannan, R. Deebalakshmi, B. G. Rajagopal, V. Jeyakrishnan, M. Abbas, and S. Kalathil, "Intelligent financial forecasting using transformers, neuro-symbolic AI, and agent-based

systems," Sci. Rep., vol. 16, Art. no. 25013, May 2026, doi: 10.1038/s41598-026-49558-0.
14. S. Kodela, "Neuro-symbolic AI for policy-aware database access control and insider threat detection in banking systems," J. Inf. Syst. Eng. Manag., vol. 9, no. 4, 2024, e-ISSN 2468-4376. [Online]. Available: https://www.jisem-journal.com/download/185_Neuro.pdf.
15. S. Li and J. Gong, "A survey on neuro-symbolic auditing: A framework for verification, traceability, and correction in high-stakes AI," Neurosymbolic Artif. Intell. J., 2025. [Online]. Available: https://neurosymbolic-ai-journal.com/system/files/nai-paper-937.pdf.
16. Z. Li, R. Song, C. Sun, W. Xu, Z. Yu, and J.-R. Wen, "Can large language models mine interpretable financial factors more effectively? A neural-symbolic factor mining agent model," in Findings Assoc. Comput. Linguistics: ACL 2024, Bangkok, Thailand, Aug. 2024, pp. 3891–3902, doi: 10.18653/v1/2024.findings-acl.233.
17. G. Zeidler, "NSAR: Hybrid neuro-symbolic AML reasoner with auditable proof trees," GitHub, Aug. 2025. [Online]. Available: https://github.com/gregorizeidler/nsar-aml-reasoner.
18. D. Arumugam, "A cognitive blockchain framework for real-time fraud anticipation in dynamic financial environments," J. African Develop., vol. 6, no. 1, pp. 77–86, 2025.
19. K. Matton, R. O. Ness, J. V. Guttag, and E. Kiciman, "Walk the talk? Measuring the faithfulness of large language model explanations," in Proc. Int. Conf. Learn. Represent. (ICLR), Singapore, Apr. 2025. [Online]. Available: https://proceedings.iclr.cc.
20. "Blockchain-anchored audit trails for neuro-symbolic urban AI: Immutable accountability for autonomous decisions in smart city infrastructure," Int. J. Artif. Intell. Data Sci. Mach. Learn., vol. 6, no. 4, 2025. [Online]. Available: https://ijaidsml.org.
21. T. Zhang et al., "A neuro-symbolic and blockchain-enhanced multi-agent framework for fair and consistent cross-regulatory audit intelligence," in Proc. Int. Conf. Digit. Soc. Intell. Comput. (ICDSC), Guangzhou, China, Nov. 2025. [Online]. Available: https://dlnext.acm.org.
22. Y.-S. Hsia, F. Yu, and J.-H. R. Jiang, "Neuro-symbolic compliance: Integrating LLMs and SMT solvers for automated financial legal analysis," in Proc. 2nd IEEE/ACM Int. Conf. AI-Powered Softw. (AIware), 2025, pp. 1–10.
23. Abdusalomov, M. Mukhiddinov, S. Makhmudov, D. Bazarova, and C. S. Oğurlu, "Bridging explainable AI and legal accountability in financial fraud detection through neural-symbolic compliance," Zenodo, 2025. DOI: 10.5281/zenodo.17485218.
24. Y. Cui, X. Han, J. Chen, X. Zhang, J. Yang, and X. Zhang, "FraudGNN-RL: A graph neural network with reinforcement learning for adaptive financial fraud detection," IEEE Open J. Comput. Soc., vol. 6, pp. 426–437, 2025. DOI: 10.1109/OJCS.2025.3543450.
25. T. Awosika, R. M. Shukla, and B. Pranggono, "Transparency and privacy: The role of explainable AI and federated learning in financial fraud detection," IEEE Access, vol. 12, pp. 64551–64560, Apr. 2024. DOI: 10.1109/ACCESS.2024.3394528.
26. "Reviewing XAI for credit risk assessment and analysis: A transparent approach to financial decision-making," in Proc. IEEE Int. Conf., Sep. 2025. [Online]. Available: https://ieeexplore.ieee.org.
27. Trembovetskyi and I. Rozlomii, "Application of transformers in financial analysis: Forecasting stock prices and trading volumes," in Proc. IEEE 5th KhPI Week Adv. Technol. (KhPIWeek), Kharkiv, Ukraine, 2024, pp. 1–5.
28. S. Johri, A. Verma, B. P. Baria, and B. G. Babu, "Neuro-symbolic finance: Integrating symbolic reasoning with deep learning for explainable credit scoring and transparent financial decision systems," in Proc. Int. Conf. Commun., Comput. Emerg. Technol. (IC3ET), 2026, pp. 475–480. [Online]. Available: https://ieeexplore.ieee.org/document/11467263.
29. Z. Dai, D. Ma, J. Tong, M. Han, J. Yang, H. Liu, H. Fei, and Q. Yang, "NSR-Boost: A neuro-symbolic residual boosting framework for industrial legacy models," in Proc. ACM SIGKDD Conf. Knowl. Discovery Data Mining (KDD), Aug. 2026. DOI: 10.1145/3770855.3818453.
30. "Starlit: Privacy-preserving federated learning to enhance financial fraud detection," in Proc. IEEE Int. Conf. Privacy, Security and Trust (PST), Dubrovnik, Croatia, Oct. 14–17, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11336591.
31. "TrustFed-HE: Trustworthy federated fraud analytics for banks with homomorphic secure aggregation and poisoning-resilient training," in Proc. IEEE Int. Conf. Big Data Security on Cloud (BigDataSecurity), Boston, MA, USA, Mar. 19–20, 2026. [Online]. Available: https://ieeexplore.ieee.org/document/11458946.
32. "FinReasoner: A lightweight LLM framework for financial forecasting," in Proc. IEEE Int. Conf. Big Data and Smart Computing (BigComp), Chengdu, China, Dec. 19–21, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/11352694.
33. "A multimodal agent for stock price trend forecasting," in Proc. IEEE Int. Conf. Data Mining (ICDM), Macau, China, Dec. 8–11, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/11401488.
34. "Designing policy-compliant counterfactual explanations for fair and transparent credit risk assessment," in Proc. IEEE Int. Conf. AI and

Finance, Warangal, India, Jul. 28–30, 2026. [Online]. Available: https://ieeexplore.ieee.org/document/11656417.
35. "Integrating explainability and fairness in credit risk prediction: A hybrid approach using TabNet, LightGBM, SHAP, and counterfactual explanations on the FICO dataset," in Proc. IEEE Int. Conf. Comput. Intell. Netw. (CINE), Ghaziabad, India, Feb. 12–13, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11468663.
36. "AI-driven secure audit trails for financial compliance using immutable blockchain logs," in Proc. IEEE Int. Conf. Blockchain and Cryptocurrency (ICBC), Dubai, United Arab Emirates, Jul. 3–4, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11292506.
37. "Designing a secure and auditable financial management system using decentralized ledger technology," in Proc. IEEE Int. Conf. Adv. Comput. Commun. Syst. (ICACCS), Indore, India, Nov. 21–22, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/11608938.
38. "A graph neural network based unified framework for financial fraud detection," in Proc. IEEE Int. Conf. Comput. Intell. Netw. (CINE), Noida, India, Nov. 19–21, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/11410799.
39. "FraudGNN: Self-supervised graph neural anomaly detection for real-time financial fraud with adversarial robustness and explainable reasoning," in Proc. IEEE Int. Conf. Artif. Intell. Data Sci., Houston, TX, USA, Feb. 18–20, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11395860.
40. "LLM-augmented explainable credit card fraud detection using XGBoost," in Proc. IEEE Int. Conf. Comput. Intell. Netw. (CINE), Coimbatore, India, Nov. 24–26, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/11383394.
41. "Fraud unmasked: Leveraging explainable AI for transparent detection systems," in Proc. IEEE Int. Conf. Adv. Comput. Commun. Syst. (ICACCS), Kolkata, India, Nov. 20–21, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11495658.
42. "Blockchain enhanced explainable artificial intelligence (XAI) for real time credit card fraud detection," in Proc. IEEE Int. Conf. Comput. Intell. Netw. (CINE), Mysore, India, Sep. 19–20, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/11411087.
43. "Responsible LLM deployment for high-stake decisions by decentralized technologies and human-AI interactions," in Proc. IEEE Int. Conf. Hum.-Mach. Syst. (ICHMS), Abu Dhabi, United Arab Emirates, May 26–28, 2025, doi: 10.1109/ICHMS65439.2025.11154208.
44. "A modular framework for decentralized explainable AI using blockchain," in Proc. Int. Conf. Artif. Intell. Speech Technol. (AIST), 2025. [Online]. Available: https://ieeexplore.ieee.org/document/11411087.
45. "Neuro-symbolic compliance-as-code: Toward explainable, adaptive, and self-evolving governance systems," in Proc. IEEE Int. Conf., 2026. [Online]. Available: https://ieeexplore.ieee.org.
46. S. M. Rahnama, "DeFiSentinel: AI-enhanced decentralized finance architecture with advanced cryptographic smart contracts for data integrity and threat resilience," IEEE Access, vol. 13, pp. 83107–83122, 2025, doi: 10.1109/ACCESS.2025.3568842.
47. "QFNN-FFD: Quantum federated neural network for financial fraud detection," in Proc. IEEE Int. Conf. Quantum Softw. (QSW), 2025, doi: 10.1109/QSW67625.2025.00015.
48. Naidu, "LLM-augmented anomaly explanation for real-time banking fraud detection: A hybrid neuro-symbolic approach," in Proc. 5th Int. Conf. Computer Networks, Big Data and IoT (ICCBI), 2026, pp. 692–698.
49. V. Mohan, "Ethics-aware hybrid graph–neuro-symbolic framework for adaptive fraud management in large-scale financial systems," IEEE Access, vol. 14, 2026.
50. R. Chilukala, Z. T. Mamadiyarov, T. Khorolskaya, S. Gururani, M. Toumasheva, and R. B. Naresh, "Explainable AI for credit scoring in cloud enterprise banking systems," in Proc. IEEE Int. Conf., Mathura, India, Sep. 2025, pp. 1–6.
51. L. Rao, M. Tsaryk, A. Benhnini, and I. Raicu, "CryptexLLM: How LLM generalizability forecasts high volatility," in Proc. IEEE Int. Conf. eScience (eScience), Chicago, IL, USA, Sep. 2025, pp. 361–362.
52. S. S. Mohanrishnan, "An information-centric blockchain and risk intelligence model for secure financial transactions in smart banking ecosystems," in Proc. 2nd Int. Conf. Intelligent Technologies for Sustainable Electric and Communications Systems (iTech SECOM), Coimbatore, India, Oct. 2025, pp. 1–6.
53. D. M. R. Putra, I. M. R. Parlambang, P. N. Manurung, A. Purwarianti, and N. P. Utama, "Multi-domain agentic fraud detection system," in Proc. IEEE Int. Conf. Industry 4.0, Bali, Indonesia, Jul. 2026, pp. 1–8.
54. G. Aminian, A. Elliott, T. Li, T. C. H. Wong, V. C. Dehon, Ł. Szpruch, C. Maple, C. Read, M. Brown, G. Reinert, and M. Mamouei, "FraudTransformer: Time-aware GPT for transaction fraud detection," in Proc. IEEE Conf. Artificial Intelligence (CAI), Granada, Spain, May 2026, pp. 850–856, doi: 10.1109/CAI68641.2026.11536152.
55. S. Goel and H. Apuri, "PolicyAgent: A multi-agent generative AI framework for automated reasoning

checks and policy-constrained compliance in financial services," in Proc. IEEE Int. Conf., 2026, pp. 1–8.
56. M. S. Khan, A. W. Khan, F. Khan, M. A. Khan, and T. K. Whangbo, "Automating regulatory compliance in Financial DevOps: An LLM-based framework for continuous audit readiness," in Proc. IEEE Int. Conf., Dubai, UAE, Jun. 2026, pp. 1–6.
57. T. Vengatesh, K. K. Kumar, K. Belliappa, M. B. Suthar, T. M. Suthar, G. B. H. Bindu, J. Bhavsar, and U. Linginedi, "Explainable AI using SHAP for transparent and interpretable decision-making in financial risk assessment and credit scoring," in Proc. IEEE Int. Conf., Dubai, UAE, Jul. 2025, pp. 1–6.
58. D. Keshavamurthy, M. Kumar, and F. Oroumchian, "An AI-based framework for secure and transparent banking: Integrating adversarial robustness, interpretability, and organizational modeling," in Proc. ACM/IEEE Int. Conf., 2025.
59. Shetty et al., "Transparent loan risk assessment through machine and deep learning," in Proc. IEEE Int. Conf., Bengaluru, India, Dec. 2025, pp. 1–6.
60. Pamisetty, S. Paleti, B. Adusupalli, J. Singireddy, R. Inala, and K. C. Nagabhyru, "Explainable AI systems for credit scoring and loan risk assessment in digital banking platforms," in Proc. 2025 IEEE 13th Int. Conf., 2025.
61. Boorugupalli, A. K. Kulkarni, AmalaSuzana, Diwakaran M, S. Ponnusamy, and S. K. S, "A multi-layered banking security framework integrating AI, biometric MFA, hybrid encryption, and blockchain auditability," ITM Web of Conferences, vol. 76, 2025, Art. no. 02002.
62. F. Yang, Z. Huang, Z. Wen, L. Ye, and Y. Qiao, "Blockchain-empowered AI for FinTech services computing: A verifiable framework for transparent and sustainable credit risk assessment," IEEE Trans. Services Computing, vol. 19, no. 3, pp. 1933–1948, 2026.
63. "AI-driven fraud detection and risk forecasting framework for real-time financial transactions," Zenodo, Mar. 2026. [Online]. Available: https://zenodo.org.
64. Y. Cui, X. Han, J. Chen, X. Zhang, J. Yang, and X. Zhang, "FraudGNN-RL: A graph neural network with reinforcement learning for adaptive financial fraud detection," IEEE Open J. Comput. Soc., vol. 6, pp. 426–437, 2025, doi: 10.1109/OJCS.2025.3543450.
65. "DriftShield: Autonomous fraud detection via actor-critic reinforcement learning with dynamic feature reweighting," IEEE Open J. Comput. Soc., vol. 6, 2025.
66. "A conceptual real-time and privacy-preserving architecture for banking fraud detection," in Proc. IEEE Int. Conf., Oct. 2025. [Online]. Available: https://ieeexplore.ieee.org.
67. "Multi-agent systems for collaborative and proactive fraud prevention in distributed AI-driven financial platforms," in Proc. IEEE Int. Conf., Nov. 2025. [Online]. Available: https://ieeexplore.ieee.org.
68. "Advancing explainable AI in deep reinforcement learning for fraud detection in financial systems," in Proc. IEEE Int. Conf., Oct. 2025. [Online]. Available: https://ieeexplore.ieee.org.
69. "Explainable machine learning frameworks for robust credit risk assessment and regulatory compliance in digital banking," in Proc. IEEE Int. Conf., May 2026. [Online]. Available: https://ieeexplore.ieee.org.
70. "Explainable AI and federated learning for privacy-preserving loan default prediction in cooperative banking networks," in Proc. IEEE Int. Conf., Mar. 2026. [Online]. Available: https://ieeexplore.ieee.org.
71. "Federated explainable AI for fair and inclusive credit scoring: A comprehensive review," in Proc. IEEE Int. Conf., 2026. [Online]. Available: https://www.semanticscholar.org.
72. "AI-driven secure audit trails for financial compliance using immutable blockchain logs," in Proc. IEEE Int. Conf., Jul. 2025. [Online]. Available: https://ieeexplore.ieee.org/document/11292506.
73. "LogProof - Ensuring integrity and immutability of audit logs using blockchain," in Proc. IEEE Int. Symp. Smart Electron. Syst. (iSES), Dec. 2025. [Online]. Available: https://www.computer.org.
74. "Large-scale financial forecasting using advanced GAI-based large language models and time series analysis," in Proc. IEEE Int. Conf., Mar. 2026. [Online]. Available: https://ieeexplore.ieee.org.
75. "A framework for quantifying ethical and regulatory risks in big data analytics and AI-assisted banking software development," in Proc. IEEE Int. Conf., Dec. 2025. [Online]. Available: https://ieeexplore.ieee.org.
76. X. Dastile and T. Celik, "Counterfactual explanations with multiple properties in credit scoring," IEEE Access, vol. 12, pp. 112345–112358, 2024, doi: 10.1109/ACCESS.2024.3441037.
77. "Counterfactual explanations of black-box machine learning models using causal discovery with applications to credit rating," in Proc. Int. Joint Conf. Neural Netw. (IJCNN), 2024, doi: 10.1109/IJCNN60899.2024.10650130.
78. "Reject inference as a noisy label detection and counterfactual correction task," in Proc. IEEE Int. Conf. Data Mining (ICDM), 2025, pp. 1–8.
79. "A constraint-aware framework for adversarial robustness in UPI fraud detection," in Proc. IEEE Int. Conf. Comput. Intell. Netw. (CINE), Theni, India, Oct. 2025, pp. 1–6.
80. "Foe for fraud: Transferable adversarial attacks in credit card fraud detection," in Proc. IEEE Int. Conf. Web Services (ICWS), 2025, pp. 286–292.
81. "Trustworthy deep learning techniques for credit risk assessment," in Proc. IEEE Int. Conf. Comput.

Intell. Netw. (CINE), Coimbatore, India, Aug. 2024, pp. 1–6.
82. "Self-evolving AI agents for financial risk prediction using continual learning and neuro-symbolic reasoning," in Proc. IEEE Int. Conf. Artif. Intell. Data Sci., 2025.
83. "Harnessing earnings reports for stock predictions: A QLoRA-enhanced LLM approach," in Proc. IEEE Int. Conf. Big Data (BigData), Hangzhou, China, Aug. 2024, pp. 1–8.
84. "Stock price prediction using LLM-based sentiment analysis," in Proc. IEEE Int. Conf. Big Data (BigData), 2024, pp. 1–8.
85. "Advanced auditing solutions through the blockchain cutting edge technology," in Proc. IEEE Int. Conf. Adv. Comput. Commun. Syst. (ICACCS), Coimbatore, India, Mar. 2024, pp. 1–6.
86. "Enhancing fraud detection in payment systems using explainable AI and deep learning techniques," in Proc. IEEE Int. Conf. Intell. Comput. Data Sci. (ICDS), Marrakech, Morocco, Oct. 2024, pp. 1–6.
87. Vaghani, Z. Gong, and M. Henke, "The role of financial digital twin in the supply chain management," in Proc. Winter Simulation Conf. (WSC), Orlando, FL, USA, Dec. 2024, pp. 2975–2986, doi: 10.1109/WSC63780.2024.10838814.
88. "Integrating rule-based approach with BERT for pharmaceutical events detection in Indian financial markets," in Proc. IEEE Int. Conf. Comput. Intell. Netw. (CINE), Chikkamagaluru, India, Jul. 2024, pp. 1–6.
89. "Graph-based machine learning approaches for fraud detection in financial networks," in Proc. IEEE Int. Conf. Comput. Intell. Netw. (CINE), Sep. 2024, pp. 1–6.
90. "AI-driven fraud detection in financial transactions with graph neural networks and anomaly detection," in Proc. IEEE Int. Conf. Comput. Intell. Netw. (CINE), Apr. 2024, pp. 1–6.
91. "Homophilic and heterophilic-aware sparse graph transformer for financial fraud detection," in Proc. IEEE Int. Conf. Comput. Intell. Netw. (CINE), Jun. 2024, pp. 1–6.
92. "SSD-GNN: Fraud detection based on spectral and spatial dual graph neural networks," in Proc. IEEE Int. Conf. High Perform. Comput. Commun. (HPCC), Dec. 2024, pp. 1–8.
93. "Research on financial fraud algorithm based on federal learning and big data technology," in Proc. IEEE Int. Conf. Sensors, Electron. Comput. Eng. (ICSECE), Aug. 2024, pp. 743–748.
94. "Security-oriented architecture for blockchain-based federated learning in the financial industry," in Proc. IEEE Int. Conf. Blockchain Cryptocurrency (ICBC), May 2024, pp. 1–6.
95. "SHAP-based interpretable models for credit default assessment using machine learning," in Proc. IEEE Int. Conf. Comput. Intell. Netw. (CINE), Aug. 2024, pp. 1–6.
96. "Explainable and fair AI: Balancing performance in financial and real estate machine learning models," IEEE Access, vol. 12, pp. 154022–154034, 2024.
97. "Effective credit risk prediction using ensemble classifiers with model explanation," in Proc. IEEE Int. Conf. Comput. Intell. Netw. (CINE), Sep. 2024, pp. 1–6.
98. "Knowledge-based credit granting system in Indonesia with artificial intelligence and XAI approach," in Proc. IEEE Int. Conf. Comput. Intell. Netw. (CINE), Jan. 2024, pp. 1–6.
99. "Big data and AI in financial risk assessment: The new scorecard approach," in Proc. IEEE Int. Conf. Inf. Technol., Big Data Artif. Intell. (ICIBA), Dec. 2024, pp. 1–6.
100. "Digital twin and IoT-enabled online learning model for real-time monitoring and control of fraudulent transaction in the banking industry," in Proc. IEEE Int. Conf. Comput. Intell. Netw. (CINE), Sep. 2024, pp. 1–6.
101. P. Chatterjee, D. Das, and D. B. Rawat, "Digital twin for credit card fraud detection: Opportunities, challenges, and fraud detection advancements," Future Gener. Comput. Syst., vol. 158, pp. 410–426, Sep. 2024
102. M. L. F. Pompeu and R. H. Filho, "Enhancing explainability in public procurement audits via neuro-symbolic learning with R-GCN and semantic loss," in Proc. IEEE 50th Annu. Comput., Softw., Appl. Conf. (COMPSAC), 2026, pp. 220–226. DOI: 10.1109/COMPSAC69091.2026.00039.
103. H. Abubakar, A. I. A. Sayed, and A. B. Karaye, "Hybrid multiobjective framework for Boolean satisfiability optimization with application to financial risk prediction," J. Appl. Math., vol. 2026, Art. no. 9145091, 2026. DOI: 10.1155/jama/9145091.
104. F. De Santis, G. Park, and F. Zanichelli, "Neuro-symbolic learning for predictive process monitoring via two-stage logic tensor networks with rule pruning," in Advances in Knowledge Discovery and Data Mining, 2026. DOI: 10.1007/978-981-92-1462-4_9.
105. Y. Liu, X.-Q. Li, and C.-L. Su, "From unstructured text to automated insights: An explainable AI approach to internal control in banking systems," Systems, vol. 14, 2026.
106. V. Sudhakar, N. Singh, and V. Kadam, "NeSy-RegReason: A neuro-symbolic AI framework for interpretable and adaptive regulatory compliance in financial services," Int. J. Comput. Inf. Syst. Ind. Manag. Appl., vol. 18, no. 23s, pp. 1882–1895, 2026. DOI: 10.70917/ijcisim-2026-5844.
107. B. Alagha and I. Ozcelik, "BEATS: Practical audit trail in blockchain systems," IEEE Access, vol. 13, pp. 109657–109669, 2025. DOI: 10.1109/ACCESS.2025.3582722.
108. Sarwosri, B. Nuralamsyah, N. R. Kurniadi, A. R. Wahed, and S. Rochimah, "E-commerce transaction fraud detection using GNN-based feature

embedding and machine learning methods," IEEE Access, vol. 14, pp. 84755–84771, 2026. DOI: 10.1109/ACCESS.2026.3699163.
109. F. K. Alarfaj and S. Shahzadi, "Enhancing fraud detection in banking with deep learning: Graph neural networks and autoencoders for real-time credit card fraud prevention," IEEE Access, vol. 13, pp. 20633–20646, 2025. DOI: 10.1109/ACCESS.2024.3466288.
110. S. Kapoor, S. Jain, P. K. Gupta, Vidushi, and S. Bhardwaj, "Integrating explainability and fairness in credit risk prediction: A hybrid approach using TabNet, LightGBM, SHAP, and counterfactual explanations on the FICO dataset," in Proc. IEEE Int. Conf. Commun. Syst. Comput. (ICCSC), 2026. DOI: 10.1109/ICCSC67078.2026.11468663.
111. V. Sampathkumar, I. P. Pathak, D. K. Ramaraj, R. Kotha, and D. P. Patel, "Federated explainable AI for fair and inclusive credit scoring: A comprehensive review," in Proc. IEEE 16th Annu. Comput. Commun. Workshop Conf. (CCWC), 2026, pp. 142–151. DOI: 10.1109/ccwc67433.2026.11393798.
112. "Explainable machine learning models for credit scoring for risk forecasting in big financial datasets," in Proc. IEEE Int. Conf. Comput. Intell. Netw. (CINE), Greater Noida, India, Feb. 2026.
113. "Beyond charts — Financial prediction with language model & time series," in Proc. IEEE Int. Conf., Boston, MA, USA, Mar. 2026.
114. "Large-scale financial forecasting using advanced GAI-based large language models and time series analysis," in Proc. IEEE Int. Conf., Boston, MA, USA, Mar. 2026.
115. R. Kumari and M. Bhatia, "Digital twin-inspired secure framework for financial theft monitoring," IEEE Internet Things J., vol. 13, no. 16, pp. 37688–37695, Aug. 2026.
116. "QuanFraud: Quantum state verification scheme for fraud detection in IoT-assisted quantum-blockchain networks," IEEE Trans. Serv. Comput., vol. 19, no. 1, pp. 616–627, Jan.–Feb. 2026.
117. "Trust without borders: A federated learning approach to privacy-preserving financial fraud detection across multi-institutional networks," in Proc. IEEE Int. Conf., Hyderabad, India, Jan. 2026.
118. "Privacy-preserving federated learning for fraud detection in distributed banking networks," IEEE Internet Things J., vol. 13, no. 15, pp. 34391–34411, Aug. 2026.
119. "Temporal-directional tree-walk graph neural network for illicit transaction detection," IEEE Trans. Comput. Soc. Syst., 2026.
120. "Exposing disguises and tracing illicit flows: Dual-view graph representation learning for money laundering detection," IEEE Trans. Comput. Soc. Syst., vol. 13, no. 2, Apr. 2026.
121. "DyHom-SSL+: Topology evolution and counterfactual semantic fusion for dynamic Bitcoin AML detection," in Proc. IEEE Int. Conf., Oct. 2026.
122. "Temporal structural feature decoupling for illicit Bitcoin transaction detection using transformer and graph attention networks," in Proc. IEEE Int. Conf., Aug. 2026.
123. "LLM-guided temporal graph neural network for illicit transaction detection in blockchain networks," in Proc. IEEE Int. Conf., Oct. 2026.
124. "Designing policy-compliant counterfactual explanations for fair and transparent credit risk assessment," in Proc. IEEE Int. Conf. AI Finance, Warangal, India, Jul. 2026.
125. "A hybrid drift-aware causal and counterfactual explainability framework for transparent loan default prediction," in Proc. IEEE Int. Conf. Comput. Sci. Netw. Technol. (CSNT), 2026. DOI: 10.1109/csnt69054.2026.11502344.
126. "Explainable credit risk assessment using graph neural networks and synthetic data augmentation," in Proc. IEEE Int. Conf., Jun. 2026.
127. "Designing an AI-based financial advisor for distressed firms: A decision support framework for actionable and accounting-consistent algorithmic recourse," IEEE Access, vol. 14, pp. 20084–20099, 2026. DOI: 10.1109/ACCESS.2026.3657172.
128. "Explainable credit risk modeling with hybrid tabular deep learning and adaptive feature routing," E&M Econ. Manag., vol. 29, no. 2, pp. 197–219, 2026. DOI: 10.15240/tul/001/2026-5-005.
129. "AI-driven secure audit trails for financial compliance using immutable blockchain logs," in Proc. IEEE Int. Conf., Jul. 2025.
130. "Follow your money: Real-time transaction tracking in blockchain trading systems," in Proc. IEEE Int. Conf., Oct. 2026.