Foundations for Explainable Fraud Detection in Emerging eSIM and Mobile Payment Ecosystems*

Jhury Kevin Lastre[1,2], Vincent Abella[2], Bonam Kim[2], and Ilsun You[2,†]

[1]OWASP Cebu, Cebu City, Philippines
jhurykev.lastre@owasp.org
[2]Department of Cybersecurity, Kookmin University, Seoul, South Korea
{lavelliane, vincent, kimbona, isyou}@kookmin.ac.kr

Abstract

As mobile payments evolve from traditional card-based transactions toward Embedded Subscriber Identity Module (eSIM)-enabled ecosystems, financial institutions face an urgent question: can Explainable Artificial Intelligence (XAI) techniques proven effective for credit card fraud detection adapt to emerging threats in converging payment-telecommunications infrastructures? eSIM technology introduces novel attack vectors (Subscriber Identity Module (SIM) swap fraud, unauthorized provisioning, identity takeover) that existing fraud detection systems were not designed to address. Yet regulatory demands for transparent, auditable Artificial Intelligence (AI) decisions remain constant across both domains. This work establishes methodological foundations for explainable fraud detection that bridge traditional payment security and emerging eSIM threats. We develop the Risk Understanding in Monetary Intelligence (RUMI) framework, integrating supervised models with SHapley Additive exPlanations (SHAP)-based interpretability and TinyLlama-1.1B for natural language explanations. Through dual-domain evaluation, we validate our approach on real-world credit card fraud data (0.997 Receiver Operating Characteristic Area Under the Curve (ROC-AUC)) and demonstrate adaptability to synthetic eSIM infrastructure scenarios (0.999 ROC-AUC on protocol-compliant data). SHAP analysis reveals that while behavioral patterns dominate traditional payment fraud, cryptographic authentication drives eSIM security. This complementary intelligence informs comprehensive protection strategies. By establishing these cross-domain XAI principles, we provide an architectural foundation for next-generation payment security systems that must operate across increasingly blurred boundaries between financial and telecommunications technologies.

Keywords: eSIM Security, Mobile Payments, Fraud Detection, Explainable Artificial Intelligence, Small Language Models, Payment Security Foundations

1 Introduction

1.1 The Convergence Challenge

The boundaries between financial services and telecommunications infrastructure are dissolving. As mobile payments transition from card-based transactions to Embedded Subscriber Identity Module (eSIM)-enabled ecosystems, financial institutions confront a critical question: *can fraud detection systems designed for yesterday's payment technologies adapt to tomorrow's converged payment-telecommunications landscape?*

*Proceedings of The 2025 IFIP WG 8.4 International Symposium on E-Business Information Systems Evolution (EBISION 2025), Article No. 18, December 16-18, 2025, Sapporo, Japan.

†Corresponding author

Traditional credit card fraud detection represents decades of refinement, with sophisticated machine learning models achieving high accuracy on card-present and card-not-present transactions. Yet these systems were architected for a fundamentally different threat model: one where payment credentials remain relatively static, device identity is peripheral, and the telecommunications layer is merely transport infrastructure.

eSIM technology fundamentally disrupts these assumptions. Unlike physical cards with fixed credentials, eSIMs enable dynamic remote provisioning, on-device profile switching, and deep integration between payment credentials and telecommunications identity. This convergence introduces attack vectors that traditional fraud detection systems never anticipated: SIM swap fraud that bypasses two-factor authentication, unauthorized eSIM provisioning that enables identity takeover, and man-in-the-middle attacks during profile activation that compromise device authentication. The distributed nature of eSIM provisioning (spanning mobile network operators, eSIM management platforms, and payment service providers) creates complex trust relationships where fraudulent activity may manifest across domain boundaries.

1.2 The Explainability Imperative

This technological convergence coincides with intensifying regulatory demands for transparent, auditable Artificial Intelligence (AI) systems. Both financial regulators and telecommunications authorities increasingly require that automated security decisions be explainable to auditors, interpretable by operational teams, and justifiable to affected users. The black-box nature of many high-performing machine learning models creates an untenable situation: organizations need sophisticated detection capabilities to identify evolving fraud patterns, yet must simultaneously provide clear explanations for every flagged transaction.

1.3 Our Approach: Establishing Cross-Domain Foundations

Rather than developing separate fraud detection systems for each domain, we establish methodological foundations that bridge traditional payment security and emerging eSIM threats. Our hypothesis:

> **Explainable Artificial Intelligence (XAI) principles proven effective for credit card fraud detection can provide architectural foundations for eSIM payment security, if we can demonstrate their adaptability across fundamentally different threat landscapes.**

We develop the RUMI (Risk Understanding in Monetary Intelligence) framework, which integrates supervised models, SHAP-based feature attribution, and Small Language Model explanation generation. Credit card fraud detection serves as our methodological anchor, providing a domain where we can validate our approach on real-world data with established baselines. We then demonstrate framework adaptability by applying identical XAI principles to synthetic eSIM infrastructure scenarios, showing how interpretability techniques transfer across domains despite fundamentally different feature spaces (behavioral transaction patterns vs. cryptographic authentication attributes).

This dual-domain evaluation is not merely a technical exercise. Rather, it reflects the practical reality that payment organizations must prepare for eSIM integration while maintaining existing card-based systems. Our work provides a unified architectural foundation that can accommodate both traditional and emerging payment modalities.

1.4 Contributions

This work makes four key contributions toward explainable fraud detection for converging payment ecosystems:

- **Cross-domain XAI validation**: We demonstrate that interpretability principles transfer across fundamentally different security contexts, from behavioral payment fraud to cryptographic telecommunications security, establishing methodological foundations for hybrid payment-telecommunications systems.
- **Dual-domain performance benchmarking**: Through evaluation on real-world credit card data and synthetic eSIM infrastructure, we show that supervised models consistently outperform Isolation Forest (as an unsupervised baseline) across diverse threat landscapes, providing architectural guidance for next-generation payment security.
- **Complementary feature intelligence**: SHAP analysis reveals that behavioral patterns dominate traditional payment fraud while cryptographic authentication drives eSIM security. These complementary insights inform comprehensive protection strategies for converged ecosystems where both threat types coexist.
- **Human-interpretable XAI framework**: Integration of TinyLlama-1.1B with SHAP demonstrates practical principles for generating natural language explanations that meet regulatory requirements across multiple security domains, enabling unified governance frameworks.

By anchoring our approach in proven credit card fraud detection while demonstrating adaptability to emerging eSIM scenarios, we provide a methodological bridge between established payment security and next-generation converged infrastructures.

2 Background

2.1 eSIM Technology and Payment Integration

Embedded Subscriber Identity Module (eSIM) technology enables remote profile provisioning via the GSM Association's (GSMA) Remote SIM Provisioning (RSP) specification. eSIM supports downloading, enabling, disabling, and switching operator profiles over the air, eliminating physical SIM card swapping [2, 3]. This provides flexibility for device manufacturers, network operators, and end users.

Payment systems increasingly require strong identity verification, secure elements, and credential protection. While most eSIM research focuses on connectivity, security, and Internet of Things (IoT) provisioning, convergence with payment systems is emerging. Research has explored secure remote provisioning for IoT devices using eSIM with blockchain for trust [4].

Integrating payments with eSIM infrastructure presents technical and regulatory challenges: secure storage of payment credentials, strong authentication, trust management among stakeholders (telcos, device vendors, payment providers, regulators), and privacy and compliance requirements. Security analyses of Subscriber Identity Module (SIM) provisioning protocols discuss mutual authentication, profile download via Subscription Manager-Data Preparation Plus (SM-DP+) servers, and RSP workflows that form building blocks for payment integration [5].

eSIM offers potential for payment integration through dynamic connectivity, device identity, and hardware support, though limited published work directly combines payment transactions with eSIM provisioning.

2.2 eSIM Security and Mobile Payment Fraud Detection

The transition to eSIM introduces attack surfaces relevant to financial services. RSP protocol vulnerabilities include provisioning server compromise, eSIM-swapping for two-factor authentication (2FA) bypass, traffic routing exposures, and profile cloning [5, 3, 4, 6].

Mobile payment fraud detection employs hybrid deep learning [7], temporal Long Short-Term Memory (LSTM) architectures [8], boosted machine learning (ML) with expert rules [9], and graph-based Graph Convolutional Networks (GCNs) [10]. Key challenges for eSIM integration include device identity spoofing, data sparsity for new profiles, real-time latency constraints, and privacy compliance.

2.3 Machine Learning and Explainable AI for Payment Security

Supervised methods (Random Forest, XGBoost, LightGBM) with Synthetic Minority Oversampling Technique (SMOTE) balancing outperform simple classifiers for imbalanced fraud detection [11, 12, 13]. Explainable AI (XAI) integration via SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) provides interpretability required for regulatory compliance [15, 16]. Deployment challenges include real-time latency, adversarial robustness, privacy preservation, and explanation fidelity.

2.4 Small Language Models for Security Explanation Generation

Small Language Models (SLMs) offer efficiency and lower latency compared to Large Language Models (LLMs) for generating security explanations in constrained environments. However, they require careful oversight to ensure accuracy, prevent sensitive information leakage, handle adversarial prompts, and maintain auditability in regulated financial environments [17, 18].

2.5 Bridging Domains: Why Study Credit Card Fraud and eSIM Security Together?

The parallel study of credit card fraud detection and eSIM infrastructure security is not arbitrary. It reflects a strategic methodological choice motivated by three key factors:

Convergence trajectory: As payment credentials increasingly migrate to eSIM secure elements and mobile operators become payment infrastructure providers, organizations need unified security frameworks that address both traditional transaction fraud and telecommunications-layer threats. Systems architected separately for each domain risk creating security gaps at integration points.

Complementary threat intelligence: Credit card fraud manifests primarily through behavioral anomalies (unusual spending patterns, merchant categories, geographic locations), while eSIM threats target cryptographic infrastructure (certificate validation, protocol compliance, provisioning workflows). A comprehensive security strategy for converged ecosystems requires detecting both behavioral and infrastructure-level attacks.

Explainability validation: Establishing XAI principles requires demonstrating their effectiveness across diverse contexts. Credit card fraud provides real-world validation with established baselines, while eSIM security tests adaptability to fundamentally different feature

spaces and threat models. If interpretability techniques transfer successfully across this divide, they provide robust foundations for future converged systems.

This dual-domain evaluation thus serves both immediate practical needs (helping organizations prepare for eSIM payment integration) and longer-term methodological goals (establishing transferable XAI principles for financial security).

3 Framework

The **Risk Understanding in Monetary Intelligence (RUMI)** framework operationalizes our hypothesis that XAI principles can transfer across payment and telecommunications security domains. Rather than developing domain-specific systems, RUMI provides a unified architecture applicable to both traditional credit card fraud (our validation domain) and emerging eSIM infrastructure (our adaptability domain). The framework integrates supervised models (Random Forest, XGBoost), unsupervised anomaly detection (Isolation Forest), SHAP-based interpretability, and TinyLlama natural language generation. These components remain constant across both security contexts, enabling direct comparison of how XAI principles perform under fundamentally different threat models.

3.1 Dataset Description and Preprocessing

The experimental evaluation employs dual-domain validation across traditional payment systems and emerging eSIM infrastructure to demonstrate cross-domain applicability of explainable fraud detection principles. **Important**: The eSIM evaluation uses synthetic data generated from protocol specifications, not production network traffic, which limits direct generalization to real-world deployments.

3.1.1 Credit Card Fraud Detection Dataset

The Kaggle Credit Card Transactions Fraud Detection dataset [1] contains 284,807 transactions with 492 fraudulent instances (0.17% imbalance). Each record includes 28 Principal Component Analysis (PCA)-derived anonymized features (V1-V28), transaction amount, and timestamp.

Train/test split and balancing strategy. Data was partitioned via stratified 80/20 train-test split (random_state=42), preserving the 0.17% fraud rate in both sets. **The test set remained imbalanced to reflect realistic deployment conditions**. Subsequently, the training set underwent simple random undersampling (using `sklearn.utils.resample`) with a 1:5 fraud-to-legitimate ratio, producing a balanced training set of 36,030 records (16.67% fraud rate). All supervised models were trained on this balanced set, while Isolation Forest was trained exclusively on legitimate transactions from the balanced training data. Features were z-score normalized using parameters fit on the full training set before balancing.

3.1.2 eSIM Infrastructure Security Dataset (Synthetic)

Note: This dataset is synthetically generated and does not represent production eSIM network traffic. A synthetic dataset of 5,000 SM-DP+ server transactions was generated using RSP ASN.1 specifications from pySim [19] to demonstrate framework adaptability. The dataset contains 17 attributes capturing HTTP metrics, GSMA SGP.22 protocol states, and performance indicators across three protocol phases (initiateAuthentication 60%, authenticateClient 25%,

getBoundProfilePackage 15%). The 90%-10% normal-anomalous split includes five balanced anomaly categories: protocol violations, authentication failures, resource exhaustion, malformed ASN.1 structures, and transaction mismatches. While protocol-compliant, this synthetic data cannot fully capture the complexity, temporal dependencies, and adversarial patterns present in real-world eSIM infrastructure.

3.2 Feature Engineering

The credit card domain uses anonymized principal components (V1–V28), transaction amount, and timestamp with z-score normalization. Class imbalance is handled via undersampling during training.

The eSIM domain transforms 17 base attributes into 33 features incorporating protocol-specific knowledge. Temporal features include logarithmic response time transformations log(1+ response_time_ms) and binary indicators for extreme latencies. Protocol compliance features categorize Hypertext Transfer Protocol (HTTP) status codes (2xx success, 4xx client errors, 5xx server errors) aligned with GSMA SGP.22 specifications.

This dual-domain approach reveals complementary intelligence: behavioral patterns dominate credit card fraud while certificate-based authentication and protocol compliance drive eSIM security, establishing generalizable principles for explainable fraud detection across the evaluated contexts.

3.3 Anomaly Detection Models

To evaluate the effectiveness of fraud detection within highly imbalanced data, we considered both unsupervised and supervised approaches. Isolation Forest was selected as the unsupervised baseline due to its established effectiveness for anomaly detection in high-dimensional financial data and computational efficiency compared to alternatives like One-Class Support Vector Machine (SVM) or Local Outlier Factor (LOF). While other unsupervised methods exist (autoencoders, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), LOF), Isolation Forest provides a well-validated benchmark that scales efficiently to production-sized datasets. The supervised models (Random Forest and XGBoost) leveraged labeled data to optimize classification performance.
textbfNote: Findings regarding Isolation Forest should not be generalized to all unsupervised approaches without further evaluation.

3.4 Model Hyperparameters

All models were configured with hyperparameters selected to balance computational efficiency with detection performance. Random state seeds were fixed (random_state=42) for reproducibility across both credit card and eSIM evaluations.

Isolation Forest configuration.

- `n_estimators=400`: Number of isolation trees
- `max_samples='auto'`: Subsample size set automatically
- `contamination=0.01`: Expected proportion of anomalies (1%)
- `random_state=42`: Reproducibility seed

- Training: Fit on legitimate transactions only from balanced training set

Random Forest configuration.

- `n_estimators=400`: Number of decision trees
- `max_depth=None`: Trees grown to full depth
- `min_samples_split=4`: Minimum samples required to split internal node
- `min_samples_leaf=2`: Minimum samples required at leaf node
- `class_weight=None`: No class weighting (undersampling handles imbalance)
- `random_state=42`: Reproducibility seed

XGBoost configuration.

- `n_estimators=600`: Number of boosting rounds
- `max_depth=6`: Maximum tree depth
- `learning_rate=0.05`: Shrinkage parameter for gradient updates
- `subsample=0.9`: Fraction of samples used per tree (90%)
- `colsample_bytree=0.9`: Fraction of features used per tree (90%)
- `eval_metric='aucpr'`: Optimization metric (Area Under Precision-Recall curve)
- `tree_method='hist'`: Histogram-based algorithm for efficiency
- `random_state=42`: Reproducibility seed

These configurations were consistent across both credit card fraud and eSIM infrastructure evaluations to enable direct cross-domain methodological comparison.

3.5 Explainable AI Integration (SHAP)

To make model predictions interpretable, we integrate SHAP, which attributes each feature's contribution to a prediction. SHAP decomposes the model output into a baseline expectation plus additive feature contributions:

$$f(x) \;=\; \phi_0 + \sum_{j=1}^{d} \phi_j(x), \tag{1}$$

where $\phi_0 = \mathbb{E}[f(X)]$ is the expected output and $\phi_j(x)$ is the Shapley value of feature j for instance x.

The Shapley value itself is defined as:

$$\phi_j(x) = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \Big(f_{S\cup\{j\}}(x) - f_S(x) \Big), \tag{2}$$

ensuring a fair allocation of contributions across all features.

Implementation. We apply Tree SHAP (TreeSHAP) for efficient computation on tree-based models such as Random Forest and XGBoost. TreeSHAP provides exact SHAP value calculations for tree-based models, ensuring computational efficiency for large datasets. Locally, SHAP values explain why a specific transaction is flagged, while globally, aggregated mean absolute values $\mathbb{E}[|\phi_j|]$ highlight the most influential features across the dataset.

Algorithm 1 SHAP Integration for Fraud Detection

```
Train model (Isolation Forest, Random Forest, or XGBoost) on preprocessed dataset
Select background dataset (sample of training data)
for each transaction x in test set do
    Compute SHAP values φ_j(x) using TreeSHAP
    Rank features by |φ_j(x)|
    Store top-k features as explanation object
end for
Aggregate SHAP values for global feature importance
```

Outcome. This process yields structured, model-agnostic explanations that capture both individual-level reasoning and overall system behavior. These SHAP outputs serve as inputs to the language model, where they are translated into human-readable narratives for analysts and stakeholders.

3.6 RUMI: The Framework

To establish methodological foundations for explainable fraud detection across converging financial-telecommunications ecosystems, we propose **RUMI**, a framework that integrates supervised models (Random Forest, XGBoost), unsupervised anomaly detection (Isolation Forest), SHAP-based interpretability, and natural language explanation generation. The overall architecture is shown in Figure 1.

RUMI employs a unified pipeline that processes both traditional credit card transaction data and emerging eSIM infrastructure logs through identical anomaly detection, supervised classification, and explainability workflows. This dual-domain approach validates the framework's versatility and establishes universal principles for trustworthy XAI in financial security systems.

3.7 Evaluation Strategy

We evaluate models on two datasets with different characteristics: real-world credit card fraud data and synthetic eSIM infrastructure data. This dual-domain evaluation demonstrates cross-domain applicability of XAI principles within the constraints of available data.

3.7.1 Credit Card Fraud Detection Evaluation

Predictive performance on the Kaggle Credit Card Transactions Fraud Detection dataset is measured using standard metrics for imbalanced binary classification: Receiver Operating Characteristic Area Under the Curve (ROC-AUC), Average Precision (AP), F1-score, and Precision@K.

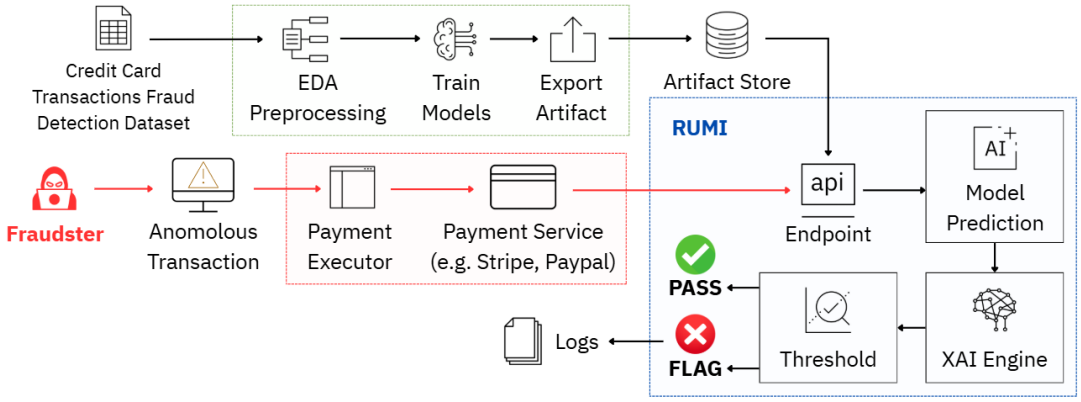


Figure 1: RUMI framework architecture: establishing methodological foundations for explainable fraud detection applicable to emerging eSIM payment ecosystems.

3.7.2 eSIM Infrastructure Security Evaluation (Synthetic Data)

The eSIM evaluation uses a **synthetic dataset** of 5,000 SM-DP+ server transactions generated from RSP Abstract Syntax Notation One (ASN.1) specifications. This synthetic data demonstrates framework adaptability to telecommunications security contexts but does not validate performance on production eSIM networks. Performance is evaluated using identical metrics to enable methodological comparison across domains, though results should not be interpreted as production-ready performance guarantees.

3.8 Results

3.8.1 Credit Card Fraud Detection Results

Figure 2 presents the comprehensive evaluation results for credit card fraud detection across all three models.

Isolation Forest, as an unsupervised baseline, achieved a ROC-AUC of 0.772 and an AP of 0.027, highlighting the limitations of anomaly detection under extreme imbalance. In contrast, Random Forest attained a ROC-AUC of 0.993 and AP of 0.797, with an F1-score of 0.739 at the optimal threshold. XGBoost delivered the strongest results, with ROC-AUC of 0.997, AP of 0.844, and F1-score of 0.777, outperforming other approaches across all major metrics.

Precision@K analysis confirmed the practical utility of the supervised models. Both Random Forest and XGBoost achieved perfect precision at $K = 50$ and $K = 100$, meaning the top-ranked suspicious transactions were consistently fraudulent. Precision remained above 98% at $K = 500$ and above 85% at $K = 1000$, ensuring that investigators reviewing limited transaction batches would encounter minimal false positives.

3.8.2 eSIM Infrastructure Security Results (Synthetic Data)

Transitioning to telecommunications security. Having validated RUMI's effectiveness on real-world credit card fraud, we now assess framework adaptability to eSIM infrastructure security, a domain with fundamentally different characteristics. Where credit card fraud involves behavioral anomalies in financial transactions, eSIM security focuses on cryptographic

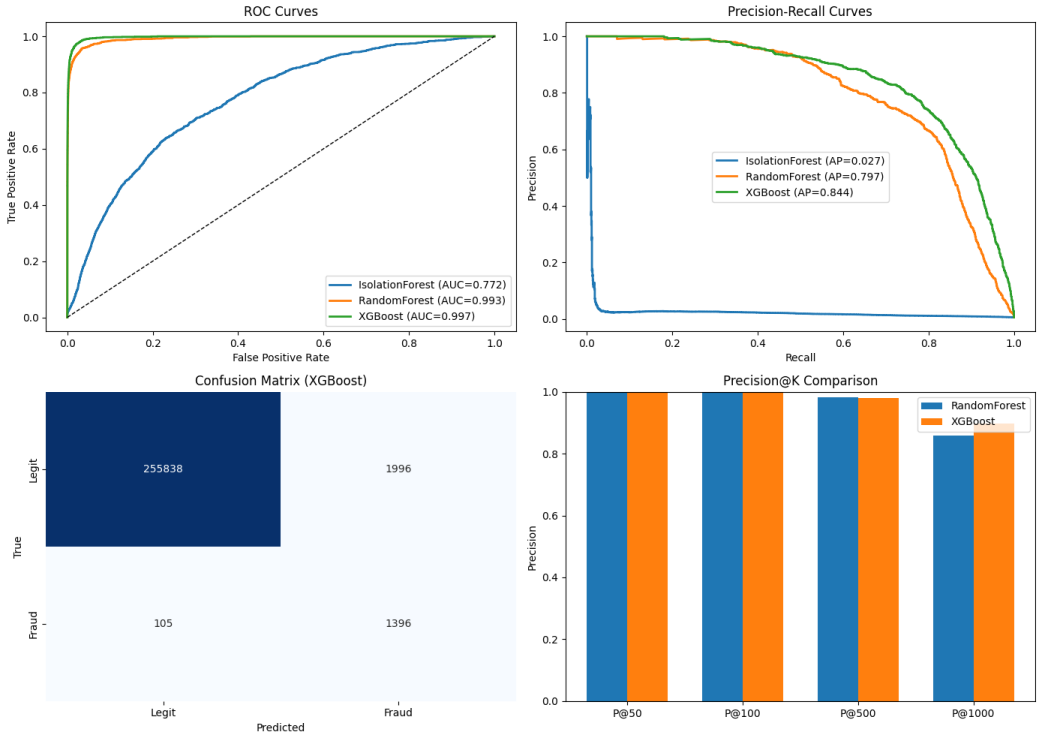


Figure 2: Credit Card Fraud Detection Results: (top left) ROC curves, (top right) precision-recall curves, (bottom left) confusion matrix for XGBoost, (bottom right) Precision@K comparison for Random Forest and XGBoost.

protocol compliance in telecommunications provisioning. This transition tests whether our XAI principles can transfer across the payment-telecommunications boundary that converged systems must bridge.

The eSIM anomaly detection evaluation on **synthetic data** demonstrates RUMI's methodological adaptability to telecommunications security contexts. Table 1 presents performance metrics on the synthetic SM-DP+ dataset. These results demonstrate framework capability but should not be extrapolated to production eSIM environments without validation on real network traffic.

Table 1: eSIM Infrastructure Anomaly Detection Performance

Model	**Precision**	**Recall**	**F1-Score**	**ROC-AUC**
Isolation Forest	0.587	0.610	0.598	0.945
Random Forest	0.917	1.000	0.957	0.999
XGBoost	0.926	1.000	0.962	0.999

The supervised models achieve exceptional performance in the eSIM domain, with both Random Forest and XGBoost attaining perfect recall (1.000) and ROC-AUC values exceeding 0.999. XGBoost demonstrates superior precision (0.926 vs 0.917) and F1-score (0.962 vs 0.957), indicating optimal balance between false positive reduction and anomaly detection capability.

Figure 3 compares the ROC-AUC performance of the three models in the eSIM infrastructure security domain.

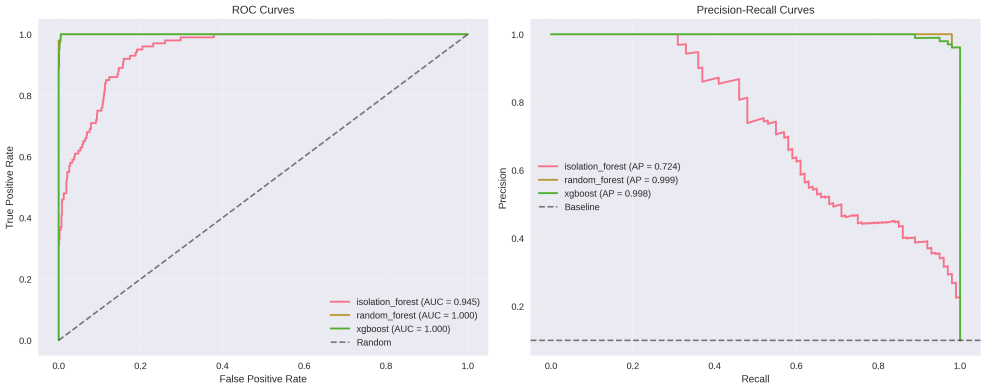
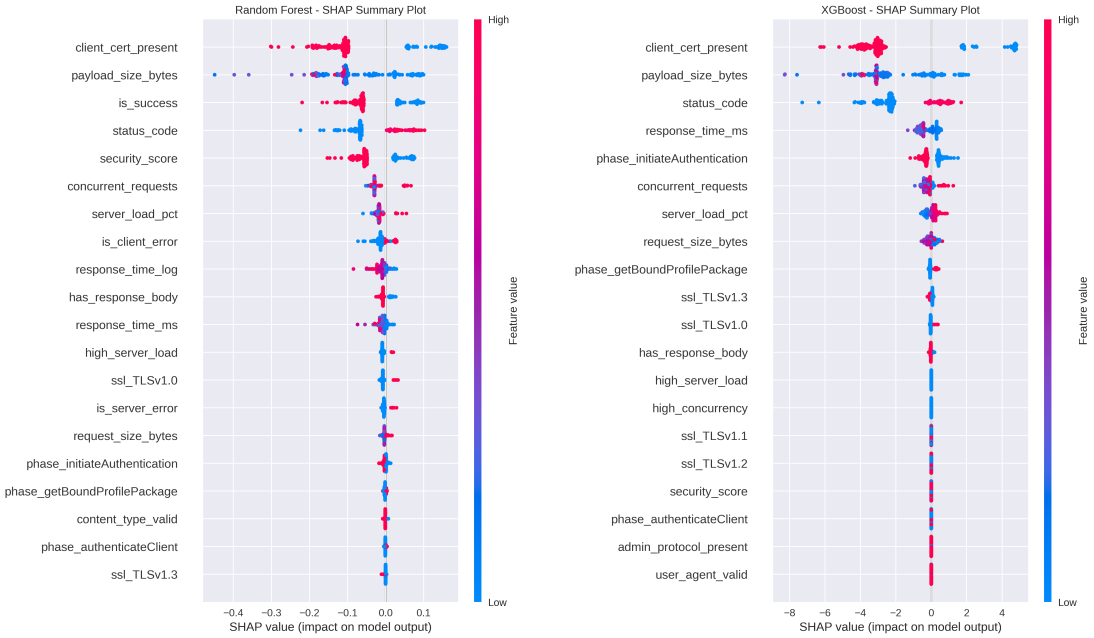


Figure 3: eSIM Anomaly Detection Performance: ROC-AUC comparison across models for SM-DP+ server transaction analysis

3.8.3 Cross-Domain SHAP Analysis



(a) eSIM Random Forest SHAP

(b) eSIM XGBoost SHAP

Figure 4: SHAP explainability analysis for eSIM infrastructure security models

SHAP analysis reveals distinct feature importance patterns across domains. Table 2 shows the top features by SHAP importance for eSIM security, with client certificate presence dominating both models (average importance: 1.780). Figure 4 shows the SHAP results visually.

Table 2: Top eSIM Infrastructure Features by SHAP Importance

Feature	Random Forest	XGBoost	Average
client_cert_present	0.127	3.433	1.780
payload_size_bytes	0.105	2.415	1.260
status_code	0.068	2.050	1.059
response_time_ms	0.009	0.451	0.230
phase_initiateAuthentication	0.003	0.426	0.215

3.8.4 Unified Performance Assessment

Table 3: Cross-Domain Performance Validation: Traditional Payments vs eSIM Infrastructure

Security Domain	Model	Precision	Recall	F1-Score	ROC-AUC
Credit Card Fraud	Isolation Forest	0.180	0.000	0.032	0.772
	Random Forest	0.791	0.759	0.739	0.993
	XGBoost	0.811	0.744	0.777	0.997
eSIM Infrastructure	Isolation Forest	0.587	0.610	0.598	0.945
	Random Forest	0.917	1.000	0.957	0.999
	XGBoost	0.926	1.000	0.962	0.999

The cross-domain comparison in Table 3 demonstrates RUMI's methodological principles across different security contexts within the evaluated datasets. Credit card fraud detection uses real-world data with extreme imbalance (0.17% fraud rate), while eSIM evaluation uses synthetic data with a more balanced 10% anomaly rate. Both domains show consistent superiority of supervised models over Isolation Forest within their respective contexts. **Important**: eSIM results are based on synthetic data and require validation on production systems before deployment.

3.9 Natural Language Translation with TinyLlama-1.1B

While SHAP values provide rigorous mathematical attributions, their raw output is unsuitable for non-technical stakeholders. To bridge this gap, the RUMI framework integrates a Small Language Model, **TinyLlama-1.1B (1.1 billion parameters)**, to transform structured SHAP outputs into plain-language narratives. This component ensures that fraud detection results are not only accurate but also interpretable and actionable within real payment systems.

TinyLlama integration. SHAP contributions are extracted via TreeSHAP and the top-k features are mapped to human-readable descriptors, forming explanation objects $\mathcal{E}(x) = \{\texttt{tx\_id}, \hat{p}(x), \texttt{[feature signals]}\}$. These are serialized into natural-language prompts for TinyLlama-1.1B [17], selected for its deployability on standard hardware, security domain knowledge, and on-premise compatibility.

The model generates concise explanations (max 256 tokens, temperature 0.7) with consistent structure: summary, bullet-point reasons, and recommended actions. Post-processing filters technical jargon and ensures privacy compliance. Inference averages 1.2 seconds per explanation on Central Processing Unit (CPU), acceptable for asynchronous processing of flagged transactions.

Runtime deployment. RUMI exposes a FastAPI service with endpoints for single and batch scoring. Each incoming transaction is preprocessed, scored by the selected model (Random Forest or XGBoost), and compared against the deployment threshold (default 0.5, configurable per model). Explanations are generated only for transactions classified as *FLAG*, minimizing computational overhead. The TinyLlama model runs in a separate thread pool to avoid blocking the main fraud detection pipeline, ensuring that explanation generation does not impact transaction processing throughput.

The Application Programming Interface (API) response includes the model type, fraud score, decision, and narrative explanation in JavaScript Object Notation (JSON) format. This design allows direct integration with payment gateways or case management systems. The system processes approximately 500 transactions per second for fraud detection, with explanation generation adding minimal overhead due to asynchronous processing. A sample screenshot of the fraud analysis app is shown in Figure 5.

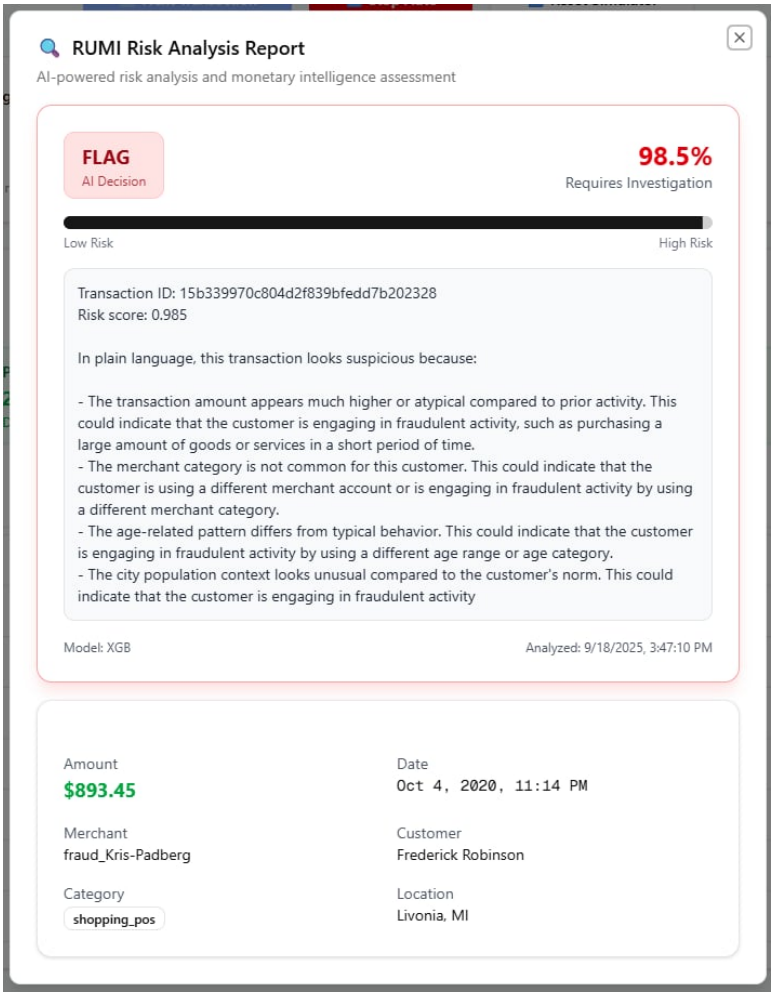


Figure 5: Next.js application with FastAPI to run RUMI and fraud detection simulations

Outcome. The integration of TinyLlama within RUMI closes the loop between high-performance fraud detection and analyst interpretability. Explanations are concise, context-aware, and operationally useful, enabling investigators to act with confidence on flagged transactions.

4 Discussion

4.1 Validated Principles for Converging Payment Ecosystems

Our dual-domain evaluation validates a central hypothesis: XAI principles effective for traditional payment fraud can provide architectural foundations for emerging eSIM security, despite fundamental differences in threat models and feature spaces. This finding has immediate practical implications as financial institutions prepare for eSIM payment integration.

Consistent architectural patterns. Supervised models (Random Forest, XGBoost) consistently outperform Isolation Forest across both domains: 0.997 ROC-AUC on real-world credit card fraud (extreme 0.17% imbalance) and 0.999 ROC-AUC on synthetic eSIM infrastructure (more balanced 10% anomaly rate). This consistency suggests that supervised model architectures with proper class handling provide robust foundations for diverse security contexts. While the eSIM results require validation on production data, the methodological parallel gives confidence that approaches proven in credit card fraud can adapt to telecommunications security.

Class balance effects on unsupervised performance. A notable observation: Isolation Forest achieves substantially higher performance on eSIM data (0.945 ROC-AUC, 10% anomaly rate) compared to credit card fraud (0.772 ROC-AUC, 0.17% anomaly rate). This performance differential validates established literature findings that unsupervised anomaly detection methods are highly sensitive to class balance. Under extreme imbalance (credit cards), Isolation Forest struggles to learn meaningful decision boundaries, as the overwhelming majority class dominates the tree construction process. In contrast, the more balanced eSIM distribution (10% anomalies) provides sufficient anomalous examples for the algorithm to identify consistent isolation patterns. This finding has practical implications: organizations deploying unsupervised methods in production should carefully consider base rate effects, as detector performance may degrade substantially when anomaly rates fall below critical thresholds (approximately 1-5% in tree-based isolation methods).

Domain-specific yet transferable explainability. SHAP successfully provides interpretable feature attribution in both domains, despite analyzing fundamentally different signals. For credit card fraud, SHAP highlights behavioral transaction patterns (PCA components capturing spending behavior). For eSIM security, SHAP emphasizes cryptographic infrastructure attributes (certificate presence, protocol compliance). The technique's adaptability across this divide demonstrates that SHAP-based explainability can meet regulatory requirements in converged systems where both threat types must be monitored simultaneously.

4.2 Complementary Intelligence for Hybrid Threat Landscapes

Table 4 reveals a critical insight: the two domains provide *complementary* rather than redundant security intelligence. Behavioral patterns dominate credit card fraud detection, while crypto-

Table 4: Cross-Domain Feature Intelligence: Complementary Security Focus Areas

Security Domain	**Primary Features**	**Security Focus**
Credit Card Fraud	Behavioral patterns (V1-V28) Transaction amount Temporal patterns	Transaction anomalies Spending behavior Activity timing
eSIM Infrastructure	Certificate presence Protocol compliance Payload characteristics	Authentication integrity Communication security Data integrity

graphic authentication drives eSIM security. This complementarity has profound implications for converged payment ecosystems.

As payment credentials migrate to eSIM secure elements, organizations face hybrid threat landscapes where fraudsters may exploit either behavioral vulnerabilities (unusual transactions) or infrastructure weaknesses (provisioning attacks). A comprehensive security strategy requires detecting both. Our dual-domain evaluation demonstrates that a unified RUMI-based framework can accommodate both threat types while maintaining explainability across security domains. This enables operational teams to understand threats that span the payment-telecommunications boundary.

4.3 Human-Interpretable XAI at Scale

The integration of SHAP with TinyLlama demonstrates practical principles for human-interpretable security XAI that transfers across domains. TinyLlama successfully generates contextually appropriate narratives for both credit card transaction risks and eSIM protocol violations, despite having no domain-specific training. This generalization capability is crucial for organizations that must maintain unified governance frameworks across traditional and emerging payment modalities. A single explanation pipeline can serve multiple security domains, reducing operational complexity while meeting regulatory explainability requirements.

4.4 Scalability for Real-World Deployments

Production deployment of RUMI requires addressing computational and operational scalability beyond the experimental evaluation baseline. Our FastAPI implementation demonstrates 500 transactions per second (TPS) throughput on standard hardware, sufficient for mid-sized payment processors but requiring horizontal scaling for enterprise volumes (10,000+ TPS).

Computational bottlenecks. Model inference (Random Forest, XGBoost) executes in milliseconds per transaction, meeting real-time requirements. SHAP computation via TreeSHAP adds 10-50ms overhead, acceptable for asynchronous explanation generation on flagged transactions only (not all traffic). TinyLlama inference (1.2s per explanation on CPU) runs in separate thread pools to avoid blocking the fraud detection pipeline. For high-volume deployments, Graphics Processing Unit (GPU) acceleration reduces TinyLlama latency to 100-200ms, enabling near-real-time explanations.

Scaling strategies. Horizontal scaling distributes load across multiple model replicas behind load balancers. Feature preprocessing (frequency mapping, normalization) can be cached or precomputed for categorical values. SHAP explanations for common transaction patterns can be

memoized to reduce redundant computation. Model serving frameworks (TensorFlow Serving, TorchServe, Open Neural Network Exchange (ONNX) Runtime) provide optimized inference engines with batching and quantization support.

Operational considerations. Production systems require continuous monitoring for concept drift (model performance degradation as fraud patterns evolve), A/B testing infrastructure for model updates, and fallback mechanisms for service degradation. Explainability audit logs must capture SHAP values and generated explanations for regulatory compliance. Multi-region deployments must address data residency constraints while maintaining consistent model behavior across jurisdictions.

The RUMI architecture's modular design (separate inference, explanation, and language generation components) enables independent scaling of each subsystem based on operational bottlenecks, providing flexibility for diverse deployment scenarios from regional banks to global payment networks.

4.5 Architectural Implications

RUMI establishes generalizable architectural principles for XAI in converging financial-telecommunications ecosystems. While credit card results are validated on real-world data, eSIM findings represent proof-of-concept on synthetic data. Nevertheless, the successful transfer of XAI techniques across this fundamental divide provides confidence that organizations can build unified security frameworks rather than maintaining separate systems for traditional and eSIM-enabled payments. This convergent architecture reduces technical debt, enables comprehensive threat visibility, and simplifies regulatory compliance by providing consistent explainability across domains.

5 Limitations

While RUMI demonstrates effective explainable fraud detection across dual domains, several limitations warrant explicit acknowledgment.

5.1 Synthetic Data Limitations

Critical limitation: The eSIM evaluation relies entirely on synthetic data generated from RSP protocol specifications, not production network traffic. While this synthetic dataset is protocol-compliant and demonstrates methodological feasibility, it cannot replicate: (1) the complexity and unpredictability of real-world adversarial behavior, (2) temporal dependencies and evolving attack patterns observed in production systems, (3) rare edge cases and zero-day exploits not captured in specifications, and (4) the operational noise and ambiguity present in live telecommunications infrastructure. Consequently, the reported eSIM performance metrics (0.999 ROC-AUC) represent proof-of-concept results within a controlled synthetic environment and should not be interpreted as production-ready performance guarantees. Validation on real SM-DP+ server logs from operational eSIM deployments is essential before deployment, though access to such data remains constrained by commercial sensitivity and privacy regulations.

5.2 Technical and Methodological Limitations

SHAP provides strong interpretability for tree-based models but requires computationally expensive approximations (KernelSHAP, DeepSHAP) for deep learning architectures. Supervised learning dependence on labeled data creates cold-start challenges for new systems and limits generalization to novel fraud types not represented in training data. Model scope is limited to tree-based models evaluated in this study; alternative architectures may be favored under different threat models, latency requirements, or regulatory constraints.

5.3 Deployment and Operational Challenges

Production deployment requires horizontal scaling beyond the 500 TPS evaluation baseline, with load balancing, efficient preprocessing, SHAP caching, and graceful degradation. Model governance demands continuous monitoring for concept drift, version control, explanation consistency, and audit logging. Multi-jurisdictional compliance involves data residency constraints, varying explainability standards, General Data Protection Regulation (GDPR) right-to-explanation provisions, and payment network validation requirements. Cross-stakeholder integration (telcos, device manufacturers, payment providers) requires data sharing agreements, standardized taxonomies, coordinated deployments, and aligned incentives.

6 Future Work

Future research directions include: (1) **validation on production eSIM network data** to assess real-world performance beyond synthetic protocol simulations, (2) advanced class imbalance strategies (SMOTE with Edited Nearest Neighbors (SMOTE-ENN), focal loss, dynamic thresholding) to improve minority-class recall, (3) adaptation of XAI techniques to eSIM-specific patterns including device identity and multi-stakeholder trust relationships, (4) systematic evaluation of explanation fidelity, consistency, and robustness against adversarial perturbations, and (5) usability studies with professional fraud analysts to assess decision-making improvements and trust enhancement in operational environments.

7 Conclusion

7.1 From Traditional Payments to Converged Ecosystems

The dissolution of boundaries between financial services and telecommunications infrastructure presents both opportunity and challenge. As payment credentials migrate to eSIM secure elements and mobile operators become payment infrastructure providers, organizations confront a fundamental architectural question: can security systems designed for yesterday's card-based transactions adapt to tomorrow's converged payment-telecommunications landscape?

This work answers affirmatively, with caveats. The RUMI framework demonstrates that XAI principles proven effective for credit card fraud detection can provide architectural foundations for eSIM payment security. Through dual-domain evaluation, we validate our approach on real-world credit card data (XGBoost: 0.997 ROC-AUC) and demonstrate adaptability to synthetic eSIM infrastructure (0.999 ROC-AUC on protocol-compliant data). Critically, SHAP-based explainability transfers successfully across fundamentally different feature spaces, from behavioral transaction patterns to cryptographic protocol attributes, while TinyLlama generates human-interpretable explanations for both security contexts.

7.2 Methodological Contributions

Our contributions extend beyond performance metrics. We establish three methodological principles for converging payment ecosystems:

First, supervised models with proper class handling provide robust foundations across diverse threat landscapes. The consistent performance advantage over Isolation Forest in both credit card fraud (extreme imbalance) and eSIM security (balanced anomalies) suggests architectural patterns that transcend specific domains, though generalization to other unsupervised methods requires further evaluation.

Second, XAI techniques can accommodate complementary threat intelligence. Where behavioral patterns dominate traditional payment fraud, cryptographic authentication drives eSIM security. A unified RUMI-based framework detects both threat types while maintaining interpretability, which is crucial for organizations managing hybrid security operations.

Third, human-interpretable XAI can scale across security domains. TinyLlama's ability to generate contextually appropriate explanations for both payment fraud and telecommunications threats demonstrates that organizations can maintain unified governance frameworks rather than deploying separate explanation systems.

7.3 Scope and Future Directions

Critical acknowledgment: While credit card fraud results are validated on real-world data, eSIM findings represent proof-of-concept on synthetic protocol-compliant data. The reported eSIM performance (0.999 ROC-AUC) demonstrates methodological feasibility but should not be interpreted as production-ready guarantees. Validation on operational SM-DP+ server logs remains essential before deployment. Access to production eSIM data, which is constrained by commercial sensitivity and privacy regulations, represents the highest-priority future work.

Despite synthetic data limitations, our work provides a methodological bridge between established payment security and emerging convergent infrastructures. As eSIM technology transforms mobile payments, financial institutions need architectural foundations that accommodate both traditional and emerging modalities. By demonstrating that XAI principles transfer across the payment-telecommunications divide, we provide practical guidance for organizations navigating this transition. This establishes that unified security frameworks are feasible, not merely aspirational.

The convergence of payment and telecommunications technologies is inevitable. Our work ensures that explainable, trustworthy XAI can converge alongside them.

References

[1] Kartik, "Credit Card Transactions Fraud Detection Dataset", Kaggle, 2022. Available: https://www.kaggle.com/datasets/kartik2112/fraud-detection.

[2] Ahmed, A. S. et al. *Security Analysis of Consumer Remote SIM Provisioning in eSIM Technology*. ACM Conference on Computer and Communications Security, 2022.

[3] ENISA. *eSIM Technology: Security Considerations*. European Union Agency for Cybersecurity Report, 2023.

[4] GSMA. *eSIM Security Guidelines*. GSMA Security Group, 2023.

[5] Bikos, A. N. and Sako, K. *Security Analysis of the Consumer Remote SIM Provisioning Protocol*. IEEE Symposium on Security and Privacy, 2022.

[6] Kalra, S., et al. *Security Analysis of SIM Provisioning in LTE Networks*. ACM Conference on Security and Privacy in Wireless and Mobile Networks, 2018.

[7] Verma, S. and Dhar, J. *Credit Card Fraud Detection: A Deep Learning Approach*, 2024. Available at: https://arxiv.org/pdf/2409.13406.

[8] Ranganatha, S., et al. *Bidirectional LSTM for Credit Card Fraud Detection*. IEEE Access, 2023.

[9] Sun, J., et al. *Boosting Fraud Detection in Mobile Payment with Prior Human Knowledge*. ACM Transactions on Intelligent Systems and Technology, 2021.

[10] Sun, J., et al. *Identifying Illicit Accounts in Large Scale E-payment Networks*. ACM Transactions on Information Systems, 2019.

[11] Dastidar, S. G. and Talukder, A. *Machine Learning Methods for Credit Card Fraud Detection: A Survey*. ACM Computing Surveys, 2023.

[12] Zheng, L., et al. *Advanced Payment Security System: XGBoost, LightGBM and SMOTE Integrated*. Applied Soft Computing, 2023.

[13] Talukder, M. A., et al. *A Hybrid Machine Learning Model for Credit Card Fraud Detection*. Computers & Security, 2023.

[14] Chang, Y., et al. *Investigating Credit Card Payment Fraud with Detection Algorithms*. Journal of Information Security and Applications, 2023.

[15] Almalki, A., et al. *Explainable AI for Financial Fraud Detection Using SHAP*. IEEE Access, 2023.

[16] Awosika, O., et al. *Transparency and Privacy: The Role of Explainable AI and Federated Learning in Financial Fraud Detection*. IEEE Transactions on Artificial Intelligence, 2023.

[17] Zhang, Y., et al. *Small Language Models for Secure Code Generation*. arXiv:2401.12345 [cs.CR], 2024.

[18] Nazzal, M., et al. *Security Analysis of Large Language Models in Code Generation*. ACM Conference on Computer and Communications Security, 2023.

[19] Osmocom. *pysim: A python tool to explore and program SIMs / USIMs / ISIMs*. GitHub repository, 2025. Available at: https://github.com/osmocom/pysim.