**ABSTRACT**

Modern banking systems use AI to detect fraudulent transactions, but often lack transparency in explaining why a transaction is flagged, leading to reduced customer trust and difficulty in meeting regulatory standards. This project proposes an Explainable Fraud Alert System that integrates behavior-based modeling, anomaly detection, and large language models (LLMs) to provide human-readable justifications for alerts. By analyzing historical transaction data, the system learns individual spending patterns and uses unsupervised learning methods to identify anomalies. These anomalies are interpreted through LLMs, which generate clear, context-aware explanations accessible via a conversational interface. A trust-based scoring mechanism ranks transaction risk to aid both customers and fraud analysts. This approach enhances user trust, supports compliance with data protection laws, and improves operational efficiency. The system represents a step toward responsible, interpretable AI in the financial sector, with future work focusing on personalization, multilingual support, and real-time scalability.