In-Context Learning for Smarter Fraud Detection in Remote Secondhand Transactions

Hyun Woo Kim
Danggeun Pay Inc.
Seoul, Republic of Korea
peter.kim@daangnpay.com

Hyun Myeong Oh
Danggeun Pay Inc.
Seoul, Republic of Korea
hammer@daangnpay.com

Sunghyon Kyeong[*]
Danggeun Pay Inc.
Seoul, Republic of Korea
devyn@daangnpay.com

Abstract

this part is not ready yet.

CCS Concepts

• Computing methodologies → Machine learning; **Natural language processing**; **• Information systems → Enterprise information systems**.

Keywords

In-context learning, context-based fraud detection, fraud detection, remote secondhand transactions

ACM Reference Format:
Hyun Woo Kim, Hyun Myeong Oh, and Sunghyon Kyeong. 2025. In-Context Learning for Smarter Fraud Detection in Remote Secondhand Transactions. In *6th ACM International Conference on AI in Finance (ICAIF '25), November 15–18, 2025, Singapore.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3604237.3626838

1 Introduction

The global market for secondhand goods has been steadily expanding, driven in large part by the rise of online platforms that facilitate non-face-to-face peer-to-peer transactions. Prominent marketplaces in this domain include Facebook Marketplace (worldwide), Danggeun Market—also known as Karrot—operating in regions such as North America, Korea, and Japan, and Mercari, which is widely used in Japan. These platforms promote the reuse of goods, contributing to environmental sustainability, and attract a growing user base who are motivated by shared social and ecological values.

To ensure secure and convenient transactions for the majority of well-intentioned users, platform providers have implemented protective measures, including escrow-based financial services. Nevertheless, the anonymity and remote nature of these platforms are frequently exploited by malicious actors. For instance, some fraudulent sellers post items at unusually low prices and fail to deliver the products, engaging in what is commonly referred to as merchant fraud. In response, platforms invest substantial effort into detecting and sanctioning such fraudulent activities.

Traditional fraud detection systems have largely relied on rule-based approaches or supervised machine learning (ML) models trained on historical transaction data. Despite ongoing model retraining enabled by MLops platforms, these approaches exhibit clear limitations in rapidly evolving environments like secondhand marketplaces, where contextual factors heavily influence transaction dynamics. In particular, the early detection of novel fraud schemes remains structurally constrained.

Recently, Large Language Models (LLMs) have emerged as a promising alternative to address these limitations. In the context of secondhand trading, LLMs can effectively analyze unstructured textual data—such as listing titles, product descriptions, and seller profiles—to identify suspicious language patterns or detect fraud strategies that resemble previously known cases. Unlike traditional models that rely solely on structured features, LLMs excel at natural language understanding and can capture subtle linguistic cues, inconsistencies in phrasing, and tone variations that might otherwise elude human analysts.

Furthermore, LLMs possess the ability to cross-reference contextual information across multiple transactions. For example, the repeated use of similar phrases, emojis, or urgent language across listings from different user accounts may be linked to a single fraud actor. This capability is significantly enhanced through In-Context Learning (ICL), which enables LLMs to perform fraud detection tasks with minimal examples and without requiring explicit model fine-tuning. ICL is particularly advantageous in scenarios where large-scale labeled datasets are unavailable and where fraud tactics evolve rapidly.

Moreover, the increasing sophistication of fraudsters—who now leverage generative AI to craft convincing phishing messages, fabricate identities, and produce deepfake documents—underscores the urgency for platforms to deploy equally advanced AI-based defense mechanisms. This technological arms race necessitates the adoption of LLM-powered, intelligent fraud detection frameworks.

In this study, we propose a novel fraud detection approach tailored to non-face-to-face secondhand trading environments, leveraging LLM-based In-Context Learning. Specifically, we extract salient features from previously confirmed fraud cases using LLMs to analyze unstructured elements such as listing titles and seller profiles. These features are then compared against ongoing transactions to assess their likelihood of being fraudulent. Finally, we evaluate the effectiveness of our approach in comparison with traditional machine learning-based detection methods to determine its potential performance gains in real-world settings.

[*]Corresponding author

ICAIF '25, Singapore

ACM ISBN 979-8-4007-0240-2/23/11
https://doi.org/10.1145/3604237.3626838

2 Related Works

2.1 ML-Based Fraud Detection

A wide range of studies in both industry and academia have sought to advance techniques for financial fraud detection. One line of research focuses on representing transaction histories between bank accounts as graphs, enabling the development of graph-based fraud detection models that significantly outperform traditional baselines in terms of F1 score performance [6, 11].

Simultaneously, increasing attention has been paid to fraud in peer-to-peer transactions within online marketplaces, where financial transactions often accompany interpersonal exchanges. A prominent example is merchant fraud, in which a scammer lists trending products at unusually low prices, receives payment, and fails to deliver the goods. This type of fraud is especially prevalent in remote secondhand platforms. Some studies have addressed this issue by analyzing fraudulent seller accounts and building machine learning-based detection models using features derived from transaction histories and product listings [1, 5].

2.2 LLM-Based Fraud Detection

With the advent of large language models (LLMs), researchers and practitioners have actively explored their potential for financial fraud detection. Traditional methods—such as logistic regression, random forests, and neural networks—have long been applied to detect fraud (e.g., in credit card transactions), but these models face limitations when dealing with highly imbalanced datasets and evolving fraud patterns [12].

Recent studies suggest that Transformer-based LLMs are better suited for capturing long-range dependencies and subtle correlations in transaction data, leading to improved detection performance [4, 8]. For example, Yu et al. (2024) demonstrate that Transformer-based models outperform conventional machine learning approaches in terms of accuracy and are particularly effective at identifying rare fraudulent cases [9, 12]. The pretraining of LLMs on vast corpora enables them to form a form of commonsense understanding of sequences, which can be further enhanced through retrieval-augmented generation (RAG) methods to boost detection capabilities [10].

Moreover, LLMs have proven useful in processing unstructured data alongside structured transactional features. Butler (2025) highlights that LLMs can detect fraud-indicative language and anomalies in textual sources such as transaction notes, emails, and chat logs. This capacity allows them to surface social engineering attempts or abnormal phrasing in online interactions—types of fraud that often evade detection by traditional rule-based or statistical systems.

2.3 In-Context Learning for Fraud Pattern Recognition

In-context learning (ICL) has emerged as a powerful paradigm that enables LLMs to perform tasks without explicit fine-tuning. Introduced by Brown et al. (2020) with the release of GPT-3, ICL allows a model to generalize to new tasks using only a prompt containing a few labeled examples [3]. This characteristic makes ICL particularly well-suited for fraud detection scenarios, where only a small number of examples of emerging fraud types may be available.

Through ICL, LLMs can implicitly learn patterns from a few in-context examples and adapt to new fraud types in real time. Liu et al. (2024) apply this concept to graph-based anomaly detection, using a handful of normal nodes as context to identify outliers in unseen graphs without additional training [7]. Similarly, Bhattacharya et al. (2025) propose a system that converts structured transaction features (e.g., amount, location, device information) into natural language descriptions and feeds them into an LLM along with a few labeled examples, enabling accurate classification of novel transactions as fraudulent or legitimate [2].

3 Methodology

this part is not ready yet.

3.1 Baseline model

this part is not ready yet.

3.2 Fine-tuning using instruct datasets

this part is not ready yet.

4 Evaluation Framework

this part is not ready yet.

4.1 ROUGE

this part is not ready yet.

4.2 RDASS

this part is not ready yet.

5 Experiments

this part is not ready yet.

5.1 KoAlpaca dataset

this part is not ready yet.

5.2 KakaoBank's CSC dataset

this part is not ready yet.

5.3 Fine-tuning training

this part is not ready yet.

- **Instruction 1**: item1
- **Instruction 2**: item2
- **Instruction 3**: item3

6 Experimental Results

this part is not ready yet.

6.1 Performance of fine-tuning models

this part is not ready yet.

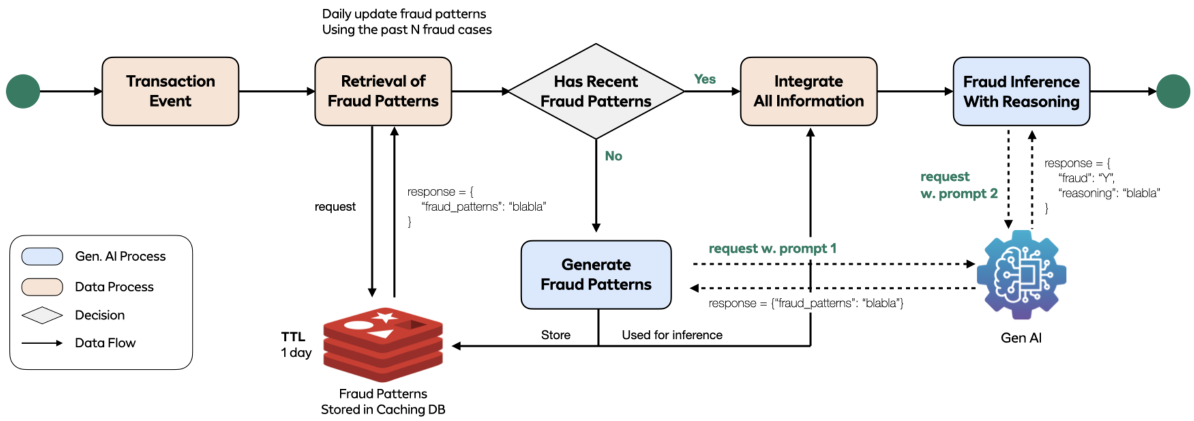

Figure 1: Experimental procedures.

6.2 Similarity measures

this part is not ready yet.

6.3 Fine-tuning with multiple instruction templates

this part is not ready yet.

7 Conclusions

this part is not ready yet.

References

[1] 2018. Detection of Fraudulent Sellers in Online Marketplaces using Support Vector Machine Approach. *International Journal of Engineering Trends and Technology (IJETT)* 57, 1 (March 2018), 48–53. doi:10.14445/22315381/IJETT-V57P210

[2] Indranil Bhattacharya and Ana Mickovic. 2024. Accounting fraud detection using contextual language learning. *International Journal of Accounting Information Systems* 53 (2024), 100682. doi:10.1016/j.accinf.2024.100682

[3] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. arXiv:2005.14165 [cs.CL] https://arxiv.org/abs/2005.14165

[4] Zhengyu Chen, Jixie Ge, Heshen Zhan, Siteng Huang, and Donglin Wang. 2021. Pareto Self-Supervised Training for Few-Shot Learning. arXiv:2104.07841 [cs.CV] https://arxiv.org/abs/2104.07841

[5] Fahim Hasan, Sourov Kumar Mondal, Md. Rayhan Kabir, Md Abdullah Al Mamun, Nur Salman Rahman, and Md. Sagar Hossen. 2022. E-commerce Merchant Fraud Detection using Machine Learning Approach. In *2022 7th International Conference on Communication and Electronics Systems (ICCES)*. 1123–1127. doi:10.1109/ICCES54183.2022.9835868

[6] Junhong Lin, Xiaojie Guo, Yada Zhu, Samuel Mitchell, Erik Altman, and Julian Shun. 2024. FraudGT: A Simple, Effective, and Efficient Graph Transformer for Financial Fraud Detection. In *Proceedings of the 5th ACM International Conference on AI in Finance* (Brooklyn, NY, USA) *(ICAIF '24)*. Association for Computing Machinery, New York, NY, USA, 292–300. doi:10.1145/3677052.3698648

[7] Shuo Liu, Di Yao, Lanting Fang, Zhetao Li, Wenbin Li, Kaiyu Feng, XiaoWen Ji, and Jingping Bi. 2024. AnomalyLLM: Few-shot Anomaly Edge Detection for Dynamic Graphs using Large Language Models. arXiv:2405.07626 [cs.LG] https://arxiv.org/abs/2405.07626

[8] Xiaopeng Liu, Yan Liu, Meng Zhang, Xianzhong Chen, and Jiangyun Li. 2019. Improving Stockline Detection of Radar Sensor Array Systems in Blast Furnaces Using a Novel Encoder–Decoder Architecture. *Sensors* 19, 16 (2019), 3470. doi:10.3390/s19163470

[9] Weimin Lyu, Songzhu Zheng, Lu Pang, Haibin Ling, and Chao Chen. 2023. Attention-Enhancing Backdoor Attacks Against BERT-based Models. arXiv:2310.14480 [cs.LG] https://arxiv.org/abs/2310.14480

[10] Anubha Pandey. 2024. Retrieval Augmented Fraud Detection. In *Proceedings of the 5th ACM International Conference on AI in Finance* (Brooklyn, NY, USA) *(ICAIF '24)*. Association for Computing Machinery, New York, NY, USA, 328–335. doi:10.1145/3677052.3698692

[11] Yeeun Yoo, Jinho Shin, and Sunghyon Kyeong. 2023. Medicare Fraud Detection Using Graph Analysis: A Comparative Study of Machine Learning and Graph Neural Networks. *IEEE Access* 11 (2023), 88278–88294. doi:10.1109/ACCESS.2023.3305962

[12] Chang Yu, Yongshun Xu, Jin Cao, Ye Zhang, Yixin Jin, and Mengran Zhu. 2024. Credit Card Fraud Detection Using Advanced Transformer Model. In *2024 IEEE International Conference on Metaverse Computing, Networking, and Applications (MetaCom)*. 343–350. doi:10.1109/MetaCom62920.2024.00064