Generative AI: Consumer Guidance for Fraudulent Credit Card Transactions

By Omkar Pardeshi, Ajay Kumar, Ethan Pollock

Project Objective & Scope

Goal: Build an end-to-end fraud detection intelligence system that improves detection accuracy and enables natural-language querying of large-scale transaction data.

Deliverables Completed:

- Baseline ML fraud models
- Synthetic fraud generation using LLM augmentation
- Fine-tuned LLM for dataset understanding
- NL → SQL → Visualization assistant
- Before/after performance comparison (FPR)
- Integrated analysis framework (DuckDB + Python)

Dataset Overview

Dataset: ~1.3M credit card transactions
Columns: Transaction timestamp, card info, merchant, category, amount, location, fraud label, etc.
Key Challenges:

- Extreme class imbalance
- High feature variety
- Realistic fraud behavior modeling

System Architecture Overview

Components:

1. Classical ML Models
2. Synthetic Data Generator
3. Fine-Tuned LLM
4. NL → SQL → Visualization Assistant
5. Metrics Comparison Engine

Baseline Classical ML Models

Models Implemented:

- RandomForestClassifier
- GradientBoostingClassifier
- LogisticRegression

Performance Summary:

Model	Baseline FPR	Baseline Accuracy	Recall	Precision
RF	0.00011247	0.9983	0.9983	0.9983
GBC	0.00045766	0.9979	0.9979	0.9978
LR	0.04665405	0.9523	0.9523	0.9933

Synthetic Fraud Generation (LLM-Augmented)

Approach:

- Identify rare fraud subpatterns
- Generate synthetic samples using fine-tuned LLM
- Maintain realistic merchant, amount, and geo distributions
- Instruction-tuned LLM outputs

Output Dataset Size Increase: 3000 rows added

Re-Training Models With Synthetic Data

Models Retrained: RF, GBC, LR

Performance After Augmentation:

Model	New FPR	Change vs Baseline	Notes
RF	0.00010472	(0.00000776)	Lower FPR. 7% performance increase.
GBC	0.00077957	(0.00032191)	Lower FPR. 70% performance increase.
LR	0.04839160	(0.00173755)	Lower FPR. 4% performance increase.

LLM Fine-Tuning on User + Transaction Profiles

Purpose:

- Teach the model spending behavior patterns
- Enable it to answer analytical questions
- Support synthetic fraud generation
- Serve as the core for NL → SQL generation

Training Approach:

- LoRA adapter
- 7000 user profiles for training
- SmolLM2-1.7B-Instruct baseline

NL → SQL → Visualization Assistant

Capabilities:

- Convert natural language into SQL
- Run optimized queries in DuckDB
- Auto-generate visualizations (line, bar, scatter)
- Supports fraud analytics questions such as:
  - “Show fraud rate by merchant category.”
  - “Plot daily fraud amounts for the last 60 days.”
  - “Compare spend velocity by state.”

Example Queries & Outputs

Examples:

1. “Top 10 categories by fraudulent transaction amount”
2. “Which states show the highest fraud concentration?”
3. “Daily legitimate spend trend in the last 90 days”

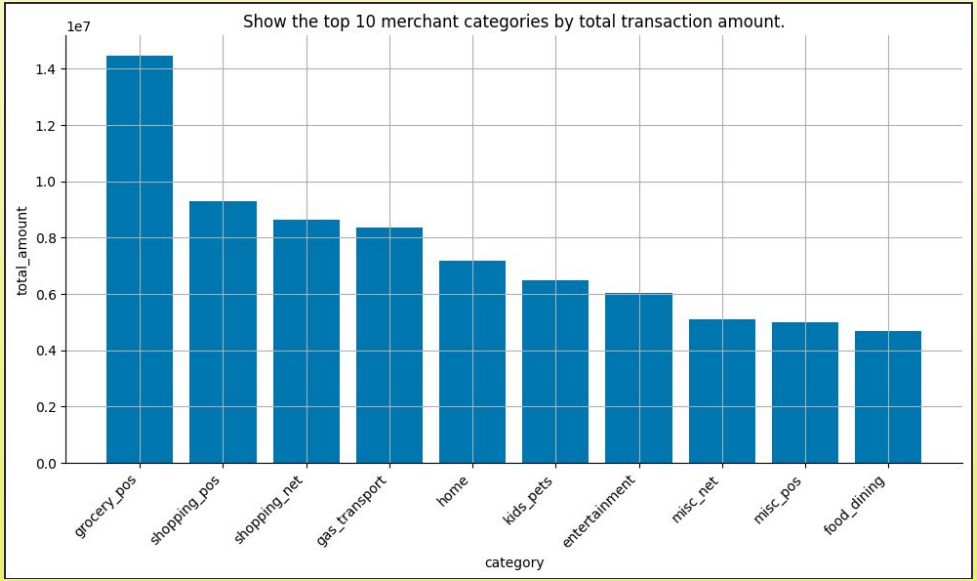
Screenshots/Charts:

```
SELECT category, SUM(amt) AS total_amount
FROM transactions
GROUP BY category
ORDER BY total_amount DESC
LIMIT 10;


---

Result preview:
```

	category	total_amount
0	grocery_pos	14460822.38
1	shopping_pos	9307993.61
2	shopping_net	8625149.68
3	gas_transport	8351732.29
4	home	7173928.11



Lessons Learned & Technical Contributions

Key Learnings:

- Synthetic fraud improves model robustness
- LLMs can safely generate SQL and augment data
- Duck DB provides extremely fast analytics at scale
- Explainability and evaluation (FPR tracking) are crucial

Contributions:

- 3-model evaluation framework
- LLM fine-tuned analytical assistant
- Synthetic fraud pipeline
- End-to-end system architecture

Conclusion & Next Steps

Completed:

✔ Baseline models
✔ Synthetic augmentation
✔ Improved FPR
✔ LLM assistant
✔ Visualization engine

References

- Tayebi, M., & El Kafhali, S. (March 17, 2025). Generative Modeling for Imbalanced Credit Card Fraud Transaction Detection. *Journal of Cybersecurity and Privacy*, *5*(1), 9. https://doi.org/10.3390/jcp5010009
- Goyal, M., & Mahmoud, Q. H. (September 4, 2024). A Systematic Review of Synthetic Data Generation Techniques Using Generative AI. *Electronics*, *13*(17), 3509. https://doi.org/10.3390/electronics13173509
- Sauber-Cole, R., Khoshgoftaar, T.M. (2022) The use of generative adversarial networks to alleviate class imbalance in tabular data: a survey. *J Big Data* 9, 98. https://doi.org/10.1186/s40537-022-00648-6
- Xu, Z., Fang, H., Han, B., Min, B., Wang, B., Zhang, S. (October 2, 2024). TagRAG: Efficient Table Retrieval and Understanding with Large Multimodal Models. *University of Wisconsin-Madison. Amazon Web Services* https://pages.cs.wisc.edu/~zxu444/home/paper/tabRAG_abs.pdf?utm_source=chatgpt.com
- Hafex, I., Hafez, A., Saleh, A., El-Mageed, A., Abohany, A. (2025). A systematic review of AI-enhanced techniques in credit card fraud detection. *Journal of Big Data*, Article number: 6. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-024-01048-8#Sec11
- Garg, R. (July 14, 2025). Generative AI in Card Fraud Detection: Benefits, Use Cases, and Industry Impacts. *Frugal Testing*. https://www.frugaltesting.com/blog/generative-ai-in-card-fraud-detection-benefits-use-cases-and-industry-impact
- Baisholan, N., Dietz, J. E., Gnatyuk, S., Turdalyuly, M., Matson, E. T., & Baisholanova, K. (March 25, 2025). FraudX AI: An Interpretable Machine Learning Framework for Credit Card Fraud Detection on Imbalanced Datasets. *Computers*, *14*(4), 120. https://doi.org/10.3390/computers14040120
- Gourav, M. (December 27, 2024). AI Analytics for Credit Card Security and Fraud Prevention. *Convin*. https://convin.ai/blog/ai-analytics-credit-card-support