Large transaction models (LTMs)

The AI breakthrough transforming fraud prevention

VISA

Contents

VISA

The AI breakthrough transforming fraud prevention

The rapid rise of Generative AI (GenAI) and Large Language Models (LLMs) has ushered in a new era of technological transformation. While LLMs have revolutionized how we interact with information, automating tasks like drafting emails, summarizing documents, generating creative content, and replacing traditional search engines, they're also being rapidly adopted by bad actors for advanced fraud.

40%

of financial institutions reporting a rise in GenAI-related attacks[1]

AI-driven cybercrime growing by over

30%

year-over-year since 2023[2]

However, by embracing the underlying principles and technology behind LLMs, the financial industry can bolster traditional fraud detection with an ability to learn, adapt, and protect at scale.

Large Transaction Models (LTMs) represent this transformative technologic leap in AI-powered fraud prevention. Just as LLMs learn the patterns and context of human language to predict and generate text, a new class of AI, LTMs, is emerging to understand the language of commerce. Applying the same principles and technology as LLMs, LTMs can help predict legitimate activity, spot anomalies, and adapt to evolving fraud tactics with unprecedented accuracy. Critically, LTMs don't replace existing fraud strategies, they supercharge them.

1 SOURCE BY" PAYMENT TRENDS UPDATE: HOW AI AND AGENTIC AI IMPACT FRAUD"
https://www.visaacceptance.com/en-us/blog/article/2025/agentic-ai-fraud-impact.html

2 SOURCE BY "SECURITY AND TRUST How AI is transforming fraud — and what you can do"
https://corporate.visa.com/en/sites/visa-perspectives/security-trust/ai-and-the-future-of-fraud.html

What is a large transaction model

With LTMs, we're not talking about your typical AI model. While most fraud models assess risk and predict outcomes, LTMs help create a foundational understanding that can be used to accelerate and amplify traditional fraud model performance.

Simply put, an LTM, or payments foundation model, is a sophisticated AI model designed to understand the "language" of commerce. Similar to LLMs trained on vast amounts of text to learn grammar and context, allowing it to predict the next word in a sentence, LTMs are generally trained on billions of anonymized transaction sequences. This transaction sequence training helps enable an LTM to predict the next transaction in a sequence and learn the intricate patterns, rhythms, and relationships that define legitimate commercial activity.

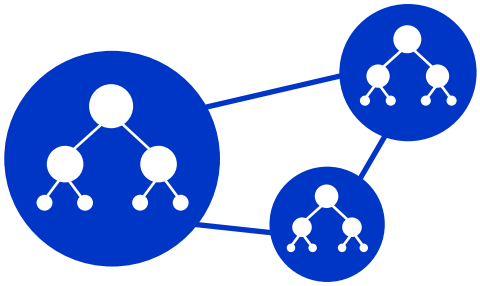
LTMs are sophisticated AI models trained on vast amounts of transaction data to predict the next transaction in a sequence, learn intricate patterns, rhythms and relationships that define legitimate commercial activity.

Transformers at work:

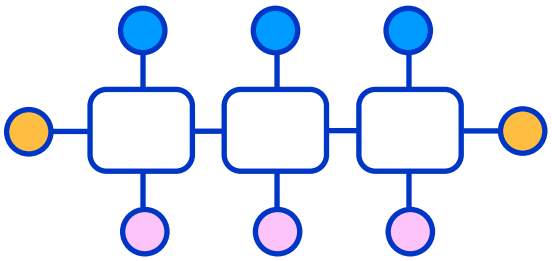
The architecture powering LTMs

To process vast amounts of transaction data and understand this "language" of commerce, LTMs utilize the same cutting-edge transformer architecture used in leading LLM. This innovative transformer architecture brings new advantages to traditional fraud prevention models commonly used today.

Machine learning has long been central to payment fraud detection. Traditional models using **Gradient Boosted Trees (GBT)** deliver efficient risk scoring by relying on predefined, engineered features from raw data. Although they depend heavily on manual feature engineering, GBTs are valued for handling non-linear relationships, managing outliers and missing values, and adapting as new threats emerge.



Deep learning represents another central, established model for fraud detection, enabling direct learning from raw, high-dimensional, sequential data like transaction histories and behavioral signals. Deep learning utilizing **recurrent neural networks (RNNs)** excel at analyzing transaction sequences by retaining prior context, uncovering evolving fraud patterns that simpler models often miss. This ability to automatically extract complex features has made deep learning a powerful tool against sophisticated fraud.



LTMs utilizing **transformers** provide new functionality compared to GBTs and RNNs. Unlike GBTs, which score transactions based on static features, and RNNs, which process sequences step by step, transformers analyze the entire transaction history at once, while avoiding the bottleneck RNNs face in processing sequences. This design enables LTMs to capture long-range dependencies efficiently and scale seamlessly with more data and parameters unlocking levels of accuracy previously unattainable. Traditional models can struggle to achieve this capability, making LTMs a true differentiator in handling massive transaction streams. The memory used by transformers is also more structured, allowing for accurate retrieval of the right information to properly contextualize current purchasing activity.

Example

Traditional models might anticipate that a traveler will book a hotel after purchasing a plane ticket. LTMs, however, going a step further by analyzing a customer's entire transaction history, can predict not only the type of hotel likely to be booked, but also additional purchases such as a luxury gift from a local boutique or personal care items that the customer tends to buy when traveling. This deeper, contextual understanding helps LTMs distinguish between genuine and suspicious activity, improving fraud detection accuracy while reducing false positives.

Transformers vs. GBTs

Strength: Transformers learn relationships automatically between raw sequential data points; GBTs rely on human-engineered features.

Benefit in fraud: Transformers can uncover hidden temporal patterns without explicit aggregation columns (e.g., "avg spend last 30 days")and then leverage the patterns to dramatically improve model performance.

Trade-off: GBTs are faster to deploy, easier to interpret, and perform well in low-data scenarios — making them popular in production banking systems.

Transformers vs. RNNs

Strength: Transformers can remove the sequential bottleneck of RNNs and scale with increasing dataset size and model complexity - much more so than RNNs whose performance plateau more quickly as dataset or model size increases.

Benefit in fraud: They can detect a suspicious transaction linked to subtle behaviors from months ago, whereas RNNs often lose signal over long spans.

Trade-off: Transformers require more compute (GPU), while RNNs can run on simpler infrastructure.

In short, while LTMs require vast amounts of high-quality data, sophisticated algorithms, and deep expertise to build effectively, when done right they deliver a step-change in fraud detection performance.

LTM embeddings and their downstream impact

While GBTs and RNNs commonly output a risk score and maybe a reason code, the core output of an LTM is a set of **transaction embeddings**. An embedding is a rich, numerical representation of a transaction that captures its deep contextual meaning. It's not just the raw data (amount, merchant, location) but a vector that encodes that transaction's relationship to the account holder's past behavior and the broader patterns of global commerce. These embeddings are the key that unlocks a more nuanced and powerful understanding of transaction risk.

Example

For example, traditional rule-based systems might identify a coffee transaction by detecting the word "coffee" in the merchant's name. In contrast, LTMs analyze a broader set of attributes, such as recurring low-value purchases in the morning combined with the merchant's name, to accurately classify coffee transactions. This approach enables the system to capture nuanced, human-like patterns beyond straightforward keyword matching.

To do this, LTMs are generally trained on vast amounts of data like LLMs, but instead of processing language, they organize and interpret financial transaction data to reveal meaningful patterns and relationships. By capturing the syntax, semantics, and pragmatics of financial behaviors, LTMs generate high-quality, contextual embeddings that reveal logical relationships and deeper meaning, similar to how grammar brings coherence to language.

LTM embeddings can provide:

Meaningful connections
Embeddings are shown to encode meaningful semantic similarities, such as clustering merchant category codes (MCCs) based on purchasing behavior.

Entity-level insight
Embeddings can summarize the transaction history of an entity like a merchant or cardholder.

Enhanced performance
Embeddings outperform hand-engineered features and other self-supervised methods on various downstream tasks, such as churn prediction, expenditure forecasting, and credit default prediction.

Interoperability
Pretrained embeddings transfer well to out-of-domain datasets, demonstrating their generalizability and effectiveness in financial modeling applications like fraud detection.

These embeddings also have various downstream impacts and may enhance a wide array of financial services including:

Fraud risk scoring
Embeddings can be fed into authorization systems in real-time to generate a highly accurate risk score, allowing for faster and more confident decisions to approve or decline a transaction.

Enhanced authentication
When a step-up authentication is required, the embedding provides a powerful additional signal to the risk model, helping to confirm the transaction with greater certainty.

Agentic commerce
In a future driven by AI agents, LTMs may help power personalized commerce by predicting and ranking top produce or service choices that an AI agent should present to the user.

The core output of a LTM is a set of transaction embeddings –rich, contextual representations that capture the relationships and patterns behind every payment. By leveraging these embeddings, financial institutions can achieve more accurate fraud prevention, stronger authentication, and deeper insights for credit scoring and personalized commerce.

Supercharging existing fraud strategies with LTMs

For the financial services industry, the most immediate impact of LTMs and associated embedding outputs is their ability to supercharge existing fraud prevention systems, helping further protect consumers and business from increasing fraud rates and false positives.

Increasing model efficiency:

The embeddings generated by LTMs can be used as a powerful new feature set in existing risk models. This has been shown to significantly increase the model's efficiency, catching more fraud while reducing the number of legitimate transactions that are incorrectly declined (false positives).

Enhanced anomaly and linkage detection:

Enhanced Anomaly and Linkage Detection: Because LTM embeddings capture subtle transaction behavioral patterns, they are exceptionally good at spotting novel fraud attacks that don't match predefined rules. They do this by organizing data so that behaviorally similar actions are co-located with one another, regardless of whether those accounts share specific entities in common. This opens the world of behavioral linkage as a powerful complement to transaction graph linkage approaches widely used in today's AML systems.

This ability also enhances the ability to link disparate, seemingly unrelated fraudulent activities across multiple accounts, revealing larger coordinated fraud rings.

LTMs will accelerate data science teams' ability to level up their model performance and realize the value of machine learning investments more quickly.

Dr. David Sutton,
Featurespace, a Visa solution,
Chief Innovation Officer

Powering the future of secure payments with Visa

Deploying LTMs requires three essential elements: a massive, diverse dataset, sophisticated infrastructure to process data, and the overall expertise to implement LTMs in real-time without adding friction. As one of the largest data sets with billions of transaction records from across the globe and decades of investment in AI infrastructure, Visa is uniquely positioned to lead this technological shift and deploy LTMs at scale.

329B+
Visa transactions[3]

4.9B+
Visa cards issued[4]

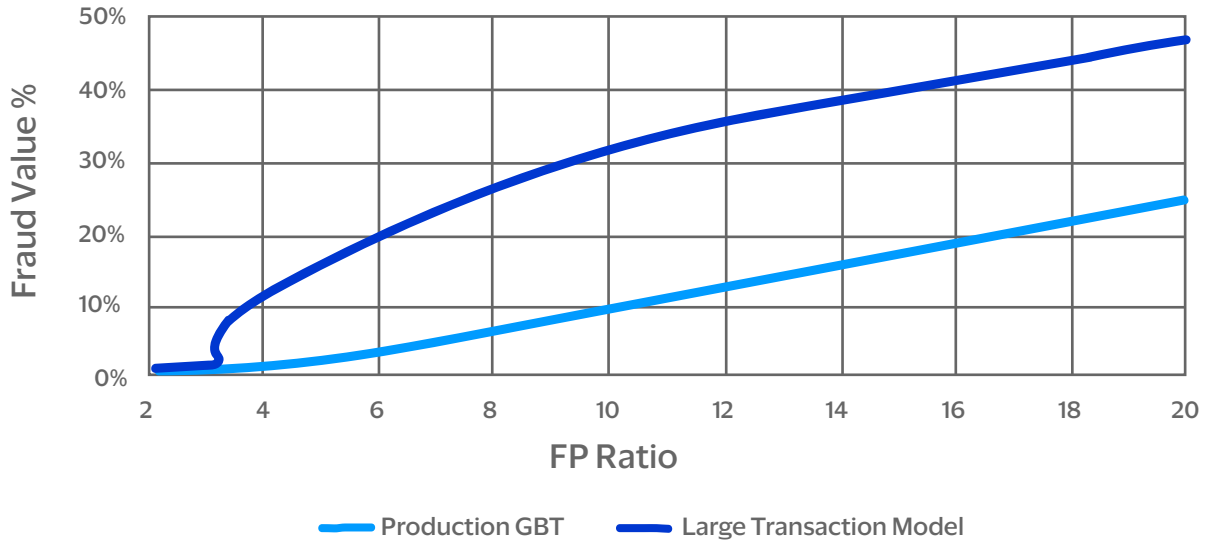
$16.7T
Total volume[3]

$3B
Invested in AI over the last 10 years[5]

3 For the 12 months ended September 30, 2025; total volume includes payments and cash volume and total transactions include 4 payments and cash transactions
4 Card data at June 30, 2025
5 30 years of AI and counting: https://corporate.visa.com/en/sites/visa-perspectives/innovation/thirty-years-of-ai-and-counting.html

5x
more fraud **caught**

And we're already seeing real results. Our proof of concept LTM, which was computed over a consortium of cards, **caught more than 5x fraud** value at a 5:1 ratio compared to a production GBT model[6]. Helping further validate that transformer-based LTMs can significantly improve fraud detection performance across market segments compared to the existing generation of fraud prevention models utilizing GBT classifiers and hand engineered features.



Learn More

To learn more about utilizing LTMs in your fraud strategy reach out to your account executive or contact sales.

Contact sales

6 Visa analysis based on aggregated VisaNet card-not-present transactions in North America using proprietary models. Results are illustrative and provided for comparative purposes only; actual performance may vary.