FinTAF

FinTech and Artificial Intelligence in Finance (FINTAF)
Issue 3

Real-Time Fraud on Real-Time Rails: An AI "Smart Circuit Breaker" for Securing Instant Payments

Senthil Nathan Dhanasekaran

Senior Banking Technologist – New Jersey, USA

Submitted 30th October 2025
Accepted 9th February 2026
Published 26th February 2026

Abstract

The launch of instant payment rails (RTP, FedNow) introduces a new, fundamental business risk: irrevocability. In this T+0, 24/7 world, the legacy T+1 security architectures that are built on static rules and batch processing are architecturally mismatched to manage this risk. Fraudulent payment is no longer an operational issue to be "fixed"; it is an immediate, irreversible financial loss.

This paper proposes that Artificial Intelligence is a viable path forward and introduces an AI-driven "smart circuit breaker" as a native component of the core payment flow, not a simple "bolt-on."

This model becomes feasible through the rich, structured data of ISO 20022, enabling real-time contextual analysis and anomaly detection before settlement. This is not just a technical upgrade; it forces boardroom-level discussions on strategic trade-offs:

- Balancing customer friction from false positives.
- Justifying the ROI on new data infrastructure.
- Solving the "explainability" (XAI) hurdle for auditors.

The paper concludes that properly architected AI is not a cost center but a critical trust enabler—and a competitive advantage for secure, reliable instant payments.

Keywords: Artificial Intelligence (AI), Instant Payments, Fraud Prevention, ISO 20022, Anomaly Detection, Authorized Push Payment (APP) Fraud, Explainable AI (XAI)

1. Introduction: The New Business Risk of 'Irrevocable, Real-Time'

1.1 Business Shift

For the last two decades, the payments industry has been driven by a simple goal: making payments faster and more convenient. With the launch of real-time rails like RTP and FedNow, the industry has largely succeeded. Businesses now have access to 24/7/365, T+0 liquidity, and customers can get instant finality. This is a massive win for business velocity, customer experience, and the entire financial ecosystem.

1.2 The New "Cost of Business"

The design of fraud detection systems introduces a new, fundamental risk into the flow - payment irrevocability - which must be recognized explicitly.

In the legacy T+1 world of ACH and wires, payment systems were architected around the luxury of time. Settlement windows, operational queues, and batch-based processing provided a critical, if unstated, business advantage - a buffer to recall, reverse, or investigate payments. If a fraudulent transaction was sent, institutions still had time to address and correct it.

In the new T+0 world, that buffer is gone. Payment is not just sent in milliseconds; it is settled, and the funds are gone. There is no recall or reversal. This transforms what was once a manageable operational issue into an instant, irreversible financial loss.

1.3 The Thesis

Traditional security and compliance models built for the T+1 world now face an architectural mismatch. The result is a significant risk exposure that legacy systems were not designed to address on their own. The thesis is simple: AI is no longer just an IT upgrade or a nice-to-have; it is the most viable path to augment controls and manage this high-speed, high-stakes risk.

2. The Obsolescence of the “T+1 Security Mindset”

To appreciate this mismatch, it is important to be candid about the current security stack. These systems are robust, but they were architected for a T+1 world and are fundamentally misaligned with the T+0 reality of instant payments. The resulting exposure stems from three core limitations. This shift also redefines operational risk-capital models, as real-time settlement collapses the window for fraud detection.

2.1 The Problem with Static Rules

For decades, the core of payment security has been the static rules engine with a certain amount of fuzzy logic baked in. This was a robust model for T+1, but in a T+0 world, it is no longer sufficient.

First, such rules are rigid and predictable given current technological advancements. Fraudsters learn to circumvent them with “low-and-slow” attacks - high-frequency, low-value payments that fly under the radar. Second, they are a primary source of business friction, generating a high rate of false positives that delay legitimate payments.

Most importantly, they are blind to modern Authorized Push Payment (APP) fraud. In an APP scam, fraud is social rather than technical - a customer is tricked into authorizing a payment. To a static rule engine, the transaction appears valid (right customer, right password) and lacks the context to detect the anomaly.

2.2 The “Batch” Fallacy

Legacy systems were built around the end-of-day (EOD) file. Institutions built large data warehouses to run analytics, generate fraud reports, and conduct Anti-Money Laundering (AML) checks in batches. This is a sound, reliable architecture for reconciliation and reporting.

For T+0 fraud prevention, however, it is a fallacy. In an irrevocable world, any security check that runs after settlement is not a prevention tool; it is a loss-reporting tool. Business needs have shifted from post-facto analysis to pre-emptive decisioning. This is a fundamental shift from batch to streaming, and many legacy systems were never designed for it. By the time the EOD report reaches management, the funds may already be realized.

2.3 The Data Limitation

A significant limitation has been the data itself. Legacy payment systems were built to parse flat files and fixed-width formats. The industry often had only a few unstructured data points—an account number, an amount, and a memo field. As the maxim goes, “garbage in, garbage out”: screening systems lacked the context needed for effective detection.

Before regulations mandated richer party information, institutions often lacked reliable, structured data on the payer, purpose, or underlying transactions. This limitation is precisely what the ISO 20022 standard is designed to solve.

3. The New Asset: ISO 20022 and AI as the New Security Stack

The shortcomings of legacy systems were defined by poor data quality. The T+0 migration brings a remedy: instant rails built on ISO 20022, a critical enabler for next-generation security.

3.1 The “New Fuel”: ISO 20022 as Rich Data Asset

Legacy programs invested heavily to extract meaning from unstructured memo fields. ISO 20022 changes the game. It is not just a new format; it is a high-quality data asset. Instead of a single memo, each payment carries a rich, structured, self-describing dataset, including:

- Structured Remittance: <RmtInf> fields with full invoice details.
- Structured Parties: <Dbtr> (Debtor) and <Cdtr> (Creditor) with discrete elements for names, addresses, and country codes.
- Purpose Codes: <Purp> fields that explicitly state the reason for the payment (e.g., salary, goods, medical).

For the first time, ecosystems can access full payment context—who is paying, who is being paid, and why. This is the fuel the industry has been missing. But fuel is not enough; the industry needs a new kind of engine to use it.

3.2 The “New Engine”: AI for Real-Time Contextual Analysis

This is where Artificial Intelligence becomes a technical necessity. Traditional rules engines are too rigid to parse this new, complex data. They can't find subtle patterns in a 2,000-character XML message. An AI model, however, can do this with the help of advanced computer power.

AI is the "engine" that can consume this new, high-quality data stream and make sense of it in real time. It moves beyond simple "if-then" logic to contextual analysis. A model can learn that a payment coded as “salary” from a known corporation to a known employee is normal, whereas the same code to a brand-new, high-risk account is anomalous. Machine-learning models can infer hidden relationships among senders, beneficiaries, and purposes, enabling pre-settlement interdiction.

This allows the model to tackle complex fraud such as APP, where AI can flag behavior that appears incorrect in context even when credentials are valid.

3.3 The "Circuit Breaker" Architectural Pattern

AI cannot be a bolt-on running in a separate batch environment. It must be architected as a "smart circuit breaker," native to the core payment flow. The flow is straightforward:

- A payment instruction (an ISO 20022 message) is initiated.
- Before being sent to the payment rail, the message is passed to the AI engine.
- The AI instantly builds a behavioral baseline, asking, "Does this payment 'look normal' for this customer, at this time, to this counterparty, for this purpose?"
- It generates a risk score in milliseconds.
- If the score is below the threshold, the payment proceeds with no delay. If the score is high, the "circuit breaker" trips, and the payment is held for review before the money leaves the institution.

This pattern provides real-time, pre-emptive fraud prevention without sacrificing instant-payment speed. It must apply to inbound credits as well, screening before posting to the beneficiary. A complete solution secures the entire, end-to-end flow.

4. The Boardroom Discussion: Balancing Innovation, Risk, and Investment

A smart circuit breaker is powerful, but implementation is not a simple IT project. It triggers strategic conversations about business models, risk appetite, and regulatory alignment. These trade-offs must be addressed at a senior level and aligned to supervisory expectations for model governance and operational resilience (e.g., OCC and BIS principles).

4.1 The New Trade-off: Customer Experience vs. Friction

The first and most immediate challenge is the "false positive." An AI model is a probabilistic system; it will make mistakes. In this context, a mistake isn't just a data error—it's a legitimate customer payment being blocked for suspected fraud.

This is no longer just an operational issue; it's a customer experience problem. If AI is tuned to be overly sensitive, it is a possibility to stop 99% of fraud but also create massive friction for 10% of good customers. If the payments systems tune it too loosely, it will let fraud through.

This is not a simple engineering decision; it's a strategic one that must be set by the business:

- What is the actual risk appetite for this new, high-speed fraud?
- What is the business cost of a false positive (customer friction, call center volume) versus a false negative (a fraudulent loss)?

- How to build a customer-service model that can instantly "unblock" a good customer who was flagged by the AI?

4.2 The Investment and ROI Question

Building this capability is not just about "buying new software." From an architectural perspective, it requires a significant investment in a new kind of data infrastructure: real-time streaming pipelines, high-speed data stores for behavioral profiles, and specialized model-serving platforms.

Traditionally, security and compliance infrastructure has been treated as a "cost center." This presents a classic boardroom challenge:

- How to build a compelling business case and secure a multi-million dollar budget for this new stack before a catastrophic, multi-million dollar fraud event makes the need obvious?
- Can payment service providers re-frame the ROI? Instead of just "losses prevented," can the provider model the value of this platform as a "business enabler" that allows the firm to confidently grow their instant payment volume?

4.3 The 'Explainability' and Compliance Hurdle

This may be the single greatest hurdle for any regulated financial institution. Many of the most powerful AI models, like deep neural networks, are "black boxes." They can provide a highly accurate risk score, but they cannot easily explain why they arrived at that score.

This is a non-starter for risk, compliance, and audit partners. "The model said so" is not an acceptable answer during a regulatory exam. This creates a fundamental technical and strategic dilemma:

- How do the payment service providers build AI models for "Explainable AI" (XAI) from day one?
- Is the business willing to accept a less accurate (but simpler and more explainable) model to satisfy audit requirements? Or do they invest heavily in the complex XAI tools needed to "interrogate" the black box?

This is a key and critical conversation that technology, risk, and compliance teams would be having before any significant investments are made in building modern security systems.

5. Conclusion: AI as a “Trust Enabler” and Competitive Advantage

For 25 years, security and compliance technology has often been viewed as a cost center. The shift to T+0, irrevocable payments now compels a fundamental re-evaluation. Rules-based gatekeeping is no longer sufficient. In a real-time world, an intelligent, instantaneous security platform is a trust enabler—providing banks, customers, and regulators the confidence to operate at T+0 speed.

AI enables the technical capabilities required, but implementation demands careful navigation of trade-offs among customer friction, investment, and regulatory explainability. The institution that solves these challenges first will not only be more secure—it will gain a competitive advantage.

By building an AI-native security stack, a bank can offer a measurably faster, safer, and more reliable payment experience, transforming a necessary technology spend into a strategic differentiator for winning and retaining the next generation of customers.

As instant payments become ubiquitous, cross-institutional data sharing and consortium-driven AI models will be critical to scaling trust beyond a single firm.

References

[1] International Organization for Standardization, *Financial services — Universal financial industry message scheme (ISO 20022:2013)*. Geneva, Switzerland: ISO, 2013.

[2] Federal Reserve Banks, *FedNow Service Operating Procedures*. New York, NY: Federal Reserve Financial Services, 2024.

[3] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach* (4th ed.). Hoboken, NJ: Pearson, 2020.