THE FRAUD ARMS RACE

How Financial Institutions Are Fighting AI-Powered Threats With AI-Powered Defenses

THOMAS SHOLA | SVP, Enterprise Technology Corporate Security Risk

THREE THREATS — THREE DIFFERENT ATTACKS

Synthetic identity: real + fabricated attributes used to create a false customer.

Deepfake ATO: cloned voice/video or altered documents used to take over a real account.

AI social engineering: scalable, personalized persuasion of customers and staff.

2026: THE TIER-ONE BANK ATTACK SURFACE

Onboarding & lending | Contact centers | Mobile and online banking

Cards & payments | Wealth and retirement | ATMs & cash-out

Enterprise AI agents and third-party integrations

26 YEARS: WHAT THE DATA ACTUALLY SHOW

FTC fraud reports: ~0.14M (2001) → 2.60M (2024)

FTC reported consumer losses: >$12.5B (2024), up 25% year over year

Nasdaq estimate: ~$485.6B global fraud losses (2023)

WHY LEGACY DEFENCES STRUGGLE

Generative scale and personalization → lower attacker cost

Synthetic media → weaker visual/voice trust

Cross-channel fraud → fragmented signals

Agentic workflows → legitimate credentials can be misused

ONE BANK, DIFFERENT DECISION CLOCKS

CARD TAP: milliseconds-to-seconds authorization; friction costs sales.

IRA / 401(k): risk-based verification and workflow over hours or days; protect life savings.

Neither has one universal regulator-mandated fraud-scoring time limit.

PRODUCTION ARCHITECTURE

CHANNEL / API → EVENT STREAM → IDENTITY + DEVICE + BEHAVIOR + GRAPH FEATURES

→ LOW-LATENCY RULES + ML SCORE → POLICY ENGINE → APPROVE / STEP-UP / HOLD

→ CASE MANAGEMENT → HUMAN REVIEW → OUTCOME FEEDBACK + MONITORING

ATMs: YES, BUT BE PRECISE

Established: skimming, card compromise, cash-out, malware / jackpotting.

AI can assist impersonation, mule recruitment and targeting upstream.

Do not equate these with proven widespread autonomous AI ATM attacks.

IAM: THE CONTROL PLANE FOR AI AGENTS

Agents can act through service identities, OAuth tokens, APIs and delegated roles.

Some agents have privileged permissions; many should not.

Prompt injection can redirect a legitimate agent to misuse authorized tools.

THE TRUSTED-AGENT PARADOX

A fraud agent is legitimately authenticated — but its instructions may be compromised.

1 Agent reads customer messages, case notes, and external intelligence.

2 Attacker embeds a prompt injection in lower-trust content.

3 Agent mistakes source content for instructions and invokes an authorized tool.

4 Fraud case, payment control, or sensitive data may be affected.

FIVE PATHS TO AGENT COMPROMISE

PROMPT INJECTION | Untrusted text redirects agent behavior.

DATA POISONING | Manipulated evidence distorts fraud scores.

TOKEN THEFT | Stolen service credentials impersonate the agent.

TOOL / API TAMPERING | Altered responses mislead its reasoning.

WORKFLOW HIJACK | Agent-to-agent steps produce unauthorized outcomes.

DEFENDING NON-HUMAN IDENTITIES

Unique agent identity • least privilege • short-lived credentials

Tool allowlists • transaction-level policy checks • separation of duties

Human approval for high-impact actions • full audit chain

LIABILITY: WHEN AI GETS IT WRONG

Missed fraud → consumer loss, restitution/disputes, supervisory exposure.

False block → customer harm, complaints, unfair treatment and lost trust.

Governance → explainability, validation, monitoring, appeals, auditability.

THE SUSTAINABLE EDGE

Not the model alone.

TRUSTED IDENTITY + CROSS-CHANNEL DATA + ACTION-LEVEL AUTHORIZATION

+ LOW-LATENCY ENGINEERING + ACCOUNTABLE HUMAN OVERSIGHT