THE NEW IDENTITY CHALLENGE

HOW AGENTIC AI IS REWRITING THE FUTURE OF FRAUD PREVENTION

experian

CONTENTS

FOREWORD

Artificial intelligence has entered a new phase of transformation. In just two years, generative AI (gen AI) has reshaped how consumers interact, how businesses operate, and how criminals exploit digital ecosystems. Now, the next evolution is here. Unlike gen AI, which creates content on demand, agentic AI systems can collaborate, plan, and act autonomously.

This shift introduces new possibilities for personalisation, automation, and operational efficiency, while simultaneously transforming the nature, speed, and sophistication of fraud and fraud prevention. Criminals are no longer simply generating fake content; they are increasingly deploying autonomous fraud agents that can operate at machine speed and scale.

As organisations move from a gen AI-driven environment to one shaped by agentic AI, the implications for fraud risk, digital identity, and the future of trusted interactions become more complex. Defences, identity strategies and operating models will be challenged in a world where both customers and attackers may be AI agents. The industry must reconsider the evolution of fraud prevention measures, historically designed for human interactions, to account for a world where identity and intent are mediated through autonomous systems.

David Britton, SVP,
Strategy and Business Development

FROM GENERATIVE TO AGENTIC AI: THE NEW DYNAMICS OF AUTOMATION AND FRAUD

Over the past two years, gen AI has transformed the global fraud landscape. **TRM Labs reports a 456% increase in gen AI–enabled scam activity from May 2024 to April 2025**[1]. This surge reflects not only the growing sophistication of generative models but also their increasing accessibility, which lowers the barrier for criminal use. As AI-generated content becomes harder to distinguish from authentic interactions, fraud is expanding exponentially across sectors and geographies.

Agentic AI represents a further step-change. These systems can plan, adapt, and execute multi-step tasks autonomously, a transformation Celent characterises as **goal-oriented, highly collaborative, context-aware orchestration with memory and tool use**[2].

Trust in an AI-powered world

Consumers are increasingly turning to AI tools to guide financial decisions. Forrester forecasts that more than half of under-50s seeking financial advice will turn to gen AI tools by 2026, reflecting a fundamental shift in how customers interpret financial information. Simultaneously, AI-powered discovery is expected to reduce human web traffic to financial services sites by 20%, while machine-initiated traffic grows by 40%[3], marking the early stages of AI-mediated customer journeys.

[1]TRM Labs, AI-enabled Fraud: How Scammers Are Exploiting Generative AI. [2]Celent, Previsory: AI Agents in Banking
[3]Forrester, Predictions October 2025: Banking and Investing

AI Explainer

Gen AI »»»»»» Agentic AI

Generative AI (gen AI) creates new content such as text, images, audio, or code. It is reactive, producing outputs only when prompted and without independent decision making.

Agentic AI builds on gen AI by adding autonomy, reasoning, memory, and the ability to use tools. It can deploy worker agents to plan and execute tasks, interact with systems, and aggregate outcomes through a core agent without ongoing human input.

In simple terms, **gen AI** makes things while **agentic AI** does things, combining generative capability with autonomous action to achieve specific goals.

THE NEXT WAVE OF FRAUD: AUTONOMOUS AND ADAPTIVE ATTACKS

As agentic AI emerges as the next wave of the AI revolution, criminals are exploiting its autonomy, adaptability, and system-level coordination to increase the scale, speed, and complexity of fraud attacks with minimal human intervention.

Deloitte forecasts that **US fraud losses could reach $40bn by 2027, up from $12.3bn in 2023**[4], driven in part by deepfake-enabled schemes and automated compromise techniques.

Emerging agentic AI threats

Autonomous and continuous fraud operations

Agentic AI can run independently, scanning for targets, generating content, creating fictitious identities or businesses at scale, testing tactics, and refining its approach in real time. This enables self-sustaining attacks, such as phishing or investment scams, that operate continuously and adapt as they progress.

Multi-step coordinated attacks

These systems can manage entire fraud workflows end-to-end, from creating synthetic identities and deepfake documents to submitting applications and navigating verification. In effect, a single agent can run an entire fraud supply chain at scale.

Hyper-personalised social engineering

With access to public or compromised data, agents can build detailed profiles and generate highly targeted scams. Real-time voice clones or deepfakes could impersonate colleagues, suppliers, or customers to authorise payments or extract sensitive information.

Tool and API exploitation

Agentic systems can interact with legitimate digital tools and APIs to automate tasks such as account creation, payment routing, or data manipulation. This means malicious activity could appear to be routine automation, thereby reducing the likelihood of detection.

Adaptive evasion of detection systems

These agents learn from blocked emails, declined transactions, or interrupted sessions, then adjust content, behaviour, or technical signals to bypass defences. Their ability to self-correct means traditional rules-based controls could quickly lose effectiveness.

Multi-agent collaboration

Future attacks may involve networks of specialised agents working together. One may probe defences, another forge documents, and another conduct social engineering. Operating collectively at machine speed, they could mirror the structure of organised fraud rings with far greater efficiency.

Agentic AI–assisted reverse engineering

LLM-powered agents rapidly analyse JavaScript to map API endpoints and validation logic (e.g., device fingerprint checks, timing gates, anti-bot challenges), adapting requests in real time to bypass defences. This compresses weeks of manual reverse engineering into minutes, enabling scalable API abuse, fingerprint spoofing, and erosion of front-end protections.

[4] *Deloitte, Deepfake Banking Fraud Risk on the Rise, 2024*

How gen AI-driven attacks are evolving

Deepfakes

An estimated 500,000 deepfakes were shared across social media platforms in 2023, and surged to 8 million in 2025[5].

Why it matters: Deepfakes now circumvent biometric checks, social-engineering safeguards, and verification flows.

Deepfakes in an agentic world

Deepfakes are primarily used in human-centric fraud, manipulating identity, trust, and authentication processes designed for people. While they remain a risk, they are less relevant in agent-to-agent interactions and are better seen as transitional tools as fraud becomes more autonomous.

Synthetic & AI-Generated Documents

Synthetic identity document fraud increased globally 195% from Q1 2024 to Q1 2025; in Europe it increased by 378%[6].

Why it matters: AI-generated documentation undermines KYC, KYB, onboarding, and lending decisions.

Synthetic & AI-Generated Documents in an agentic world

While gen AI scales the creation of fraudulent artefacts, agentic AI enables criminals to operationalise them through autonomous agents that execute attacks directly against businesses.

An estimated **500,000** deepfakes were shared across social media platforms in 2023, and surged to **8 million** in 2025[5]

[5] *https://deepstrike.io/blog/deepfake-statistics-2025*
[6] *https://financialit.net/news/artificial-intelligence/new-data-fraudsters-turning-eu-and-traditional-forgery-ai-generated*

Agentic AI examples: Fraud prevention versus fraud perpetration

Non-malicious

- Onboarding & Authorisation Agents
- Fraud & Monitoring Agents
- Compliance, Governance & Audit Agents
- Operations & Resolution Agents
- Customer Experience Agents
- Performance & Analytics Agents

Malicious

- Intrusion & Malware Agents
- Social Engineering, Phishing & Manipulation Agents
- Impersonation Agents
- Market & Financial Abuse Agents
- LLM & Application Exploitation Agents

How agentic AI can be used in fraud prevention

Agentic AI possesses three core capabilities – adaptability, autonomy, and proactivity – that distinguish it from traditional fraud prevention tools. It can respond dynamically to evolving fraud patterns, operate without constant human oversight, and anticipate emerging threats before they materialise. In a rapidly changing fraud environment, this level of intelligence and initiative can be critical to staying ahead of attackers.

It can also streamline investigative workflows by automating repetitive, time-intensive tasks. Agents can consolidate case data, generate policy-aligned reports, provide contextual transaction summaries, enhance watchlist screening, uncover hidden relationships and beneficial ownership, and automate open-source intelligence gathering. By accelerating triage and reducing manual effort, agentic AI can improve accuracy and enable fraud and compliance teams to focus on higher-value decisions.

THE RISE OF AGENTIC INTERACTIONS: AGENTIC COMMERCE AND AUTONOMOUS TRANSACTIONS

Agentic commerce is reshaping digital transactions by introducing autonomous AI agents as active participants in online shopping. Gartner® predicts that "by 2030, 20% of monetary transactions will be programmable to include terms and conditions of use, giving AI agents economic agency."[7] In addition, Celent expects that **by 2035 nearly 18% of European e-commerce value will be initiated by AI agents**[8] as consumers delegate shopping, payments, renewals, and comparisons to autonomous systems.

This evolution creates a new requirement for identity. Traditional bot-detection signals struggle to distinguish between authorised agents acting on behalf of customers and malicious agents impersonating them. As agentic commerce accelerates, the traditional four-party model of cardholder, merchant, issuer, and acquirer comes under pressure. This leaves merchants and payment systems uncertain about whether a transaction has been initiated by a human or an agent, and if that agent is genuinely authorised to act on the user's behalf. Agents may also be hijacked or manipulated, approving transactions at scale as a result of prompt injection or other forms of deception.

This introduces a critical question: **who is liable when a chargeback occurs after either authorised or unauthorised activity via an AI agent?**

Today, liability in agentic commerce largely remains with the merchant, except when 3D Secure is applied, in which case liability shifts to the credit card issuer. This highlights the need for a shared, agentic source of trust across the payments network.

> Agentic commerce also introduces new risks for consumers and for acquiring banks, particularly through the emergence of fake or synthetic merchant storefronts. These storefronts may appear operationally legitimate to both agents and payment systems, enabling fraudulent transactions, fulfilment failures, or abuse of pricing. As agents transact autonomously, traditional signals used to validate merchant legitimacy and commercial intent become less effective, increasing the risk of systemic exposure rather than isolated fraud events.

David Britton, SVP, Strategy and Business Development

[7]Gartner, Gartner's Top Strategic Predictions for 2026 and Beyond, Daryl Plummer, Alex Curry, et al., 31 October 2025
[8]Celent, Agentic Commerce Forecasts

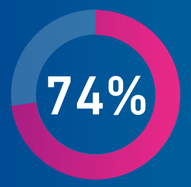
MODERN AUTHENTICATION CHALLENGES: HOW TO IDENTIFY AN AUTHORISED AI AGENT

As fraud and identity signals become harder to interpret in an environment defined by expanding data sources, automation, and real-time payments, detection frameworks are coming under increasing strain.

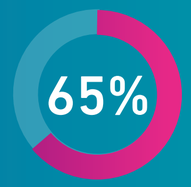
According to Gartner, "Inadequate identity controls will likely contribute to agent abuse accounting for **25%** of security breaches by 2028.[9]"

The expansion of data volume and variety, as financial institutions look beyond traditional bureau and transactional inputs to improve coverage, resilience, and decisioning speed, further increases the complexity of signal interpretation.

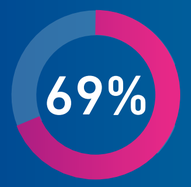
80% of institutions expect to shift toward alternative data within five years, with top use cases including[10]:



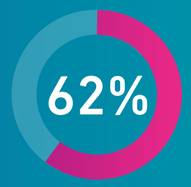
enhanced risk assessment



operational efficiency



real-time credit decisioning



fraud prevention

This challenge is further amplified when assessing bot activity, as the behaviours of legitimate and malicious automated agents are difficult to distinguish. As financial institutions ingest high-velocity behavioural, device and network-level data to distinguish human users from automation, they must also recognise that modern bot frameworks, whether used for customer service, user engagement, operational efficiency or coordinated attacks, often generate near-identical interaction patterns.

Existing CDN and bot-mitigation controls often block legitimate agent traffic due to limited visibility into whether an agent is authorised.

Instant payments have also opened new avenues for fraud, prompting regulators to enforce stronger controls against scams, new reimbursement standards, and changes to authentication. IDC notes that real-time payment schemes like FedNow, Faster Payments, RTP and SCT Inst are reshaping liability, monitoring and customer-intervention expectations[11].

[9]Gartner, How to Enable Agentic AI via API-Based Integration, Adrian Leow, Mark O'Neill, et al., 10 January 2026
[10]Future of Underwriting, Experian, 2025. [11]IDC, Worldwide Banking ERM & Compliance Technology Trends, 2026

In the US, new NACHA rules take effect this year that require both sending and receiving financial institutions to implement risk-based fraud controls. Alongside the UK's APP reimbursement model, introduced in October 2024, this signals a shift towards shared liability and reinforces the need for real-time monitoring, authorisation and intervention as agents increasingly initiate payments on behalf of customers.

Challenges distinguishing between legitimate agents and malicious bots

In today's agentic AI landscape, both legitimate and malicious bots exhibit remarkably similar behaviours, making detection and classification increasingly difficult. These include:

Use of headless browsers: Both types operate without a graphical interface, allowing for high-speed submission of forms and access to APIs.

Direct API access: Bots interact with backend systems programmatically, bypassing traditional user interfaces.

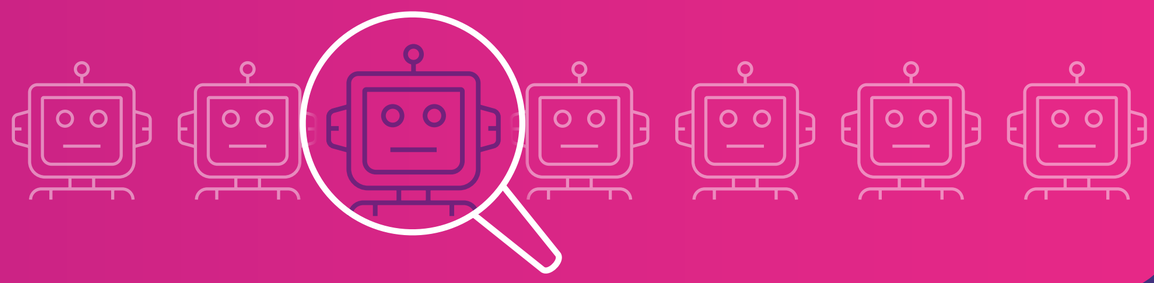
Rapid, sequential actions: Activities such as account access, data retrieval, and form submissions occur in milliseconds, mimicking the patterns of automation.

Absence of human interaction signals: No mouse movement, scrolling, or hesitation—hallmarks of human behaviour—are present.

Generative AI assistance: Attackers and enterprises alike use AI to generate realistic data, whether for customer service or synthetic identity fraud.

Fingerprint evasion techniques: Both may use anti-detect browsers, fingerprint randomisation, and residential proxy pools to obscure their identity.

Header spoofing and signature overlap: Legitimate bots may identify themselves via user-agent headers (e.g., "GPTBot"), but attackers can spoof these or use open-source agents that lack clear identifiers.



These pressures are driving a shift away from legacy, rules-based detection, which is prone to false positives and analyst overload, toward consolidated, AI-enabled orchestration and analytics platforms that unify fraud, identity, AML, and credit risk.

Methods to identify and authenticate agentic interactions must evolve to allow trusted agent activity to pass through, supported by a more reliable framework that links agent activity to human intent.

THE EVOLUTION OF DIGITAL IDENTITY: KNOW-YOUR-CUSTOMER (KYC) AND KNOW-YOUR-AGENT (KYA)

New demands on identification mean that digital identity frameworks are undergoing a transition. IDC predicts that by 2028, 70% of financial institutions will replace static credentials with continuous, context-aware identity validation for high-risk interactions.[12] At the same time, instant payment schemes such as FedNow and RTP in the United States, Faster Payments in the United Kingdom, SEPA Instant Payments (SCT Inst), and PSD3 and PSR compliant transfers in the European Union and the entire EEA are reshaping liability and reimbursement rules.[13]

These developments mean identity must become dynamic, continuous, contextual, and capable of recognising AI mediation across channels and touchpoints. The traditional model of determining 'who is on the other side' is evolving, and financial institutions must adapt accordingly.

A consensus is emerging around the need to securely establish

This points to the need for a more trustworthy framework that can link agents to the humans they represent, while capturing and preserving intent across the agentic flow.

Know-Your-Agent (KYA) is a framework to:

- Register AI agents within identity systems
- Link them to human or business owners
- Define their permissions
- Continuously monitor behaviour for anomalies

Gartner notes that, "By 2027, over **50%** of AI agents deployed in enterprises will rely on standardized frameworks like MCP or the Agent2Agent (A2A) protocol for secure, cross-system interoperability."[14]

"The momentum for agentic AI is strong, but we need transparency, data security and protection, controls and guardrails, as well as advanced monitoring before we see broad adoption of agentic systems."[15]

[12]IDC FutureScape: Worldwide Banking and Payments 2026 Predictions, #US53859825, October 2025
[13]IDC Perspective, Worldwide Banking Enterprise Risk Management and Compliance Technology Trends, 2026, #US53829225, November 2025
[14]Gartner, How to Enable Agentic AI via API-Based Integration, Adrian Leow, Mark O'Neill, et al., 10 January 2026 *[15]Forrester, Unlock AI Agent Potential With Smarter Use Case Decisions, p. 12, February 2026*

SECURING THE AGENTIC ECOSYSTEM: ESTABLISHING A NEW TRUST ARCHITECTURE

As agentic AI reshapes how digital transactions are initiated and executed, organisations must rethink their fraud prevention and identity management strategies. With agents increasingly acting on behalf of both consumers and fraudsters, the way businesses leverage trust signals such as device intelligence, behavioural biometrics and network indicators must evolve to remain effective.

Beyond commerce-specific use cases, agentic AI is also transforming the wider fraud landscape. Criminal networks can deploy autonomous agents to orchestrate attacks at scale, coordinating synthetic identities, testing stolen credentials, adapting to controls in real time and executing transactions at machine speed. The same agentic capabilities can also strengthen defence: augmenting investigator decision-making, reducing false positives through richer contextual analysis, and accelerating case handling and response.

For organisations operating in agentic commerce, these shifts fundamentally change how trust must be established at the point of interaction. New approaches are required to establish confidence in the link between a human and the agent representing them, providing richer context and enabling more accurate decisions in every agentic interaction.

Human-to-Agent (H2A) binding
A trusted link between a verified human or business and the agent acting on their behalf that captures intent, mandate, and accountability.

An agentic trust registry
A system of record that registers agents and binds them to a verified human or business, enabling real-time validation of provenance, declared capabilities and authorised scope.

Continuous trust scoring
Trust must be dynamic. Agents should be monitored and scored continuously across the ecosystem, using behavioural, transactional and network context to adapt assessments as activity evolves.

This architecture is reinforced by agentic ID graphing, extending traditional identity graphs to connect human identities, agent identifiers, behavioural patterns and device intelligence. The result is richer context for fraud risk engines and more precise decisioning across complex relationships.

At the same time, agentic AI is embedded within the control environment itself. Agent-based risk engines can generate adaptive scoring logic, respond to emerging attacks and optimise outcomes in real time. Continuous model risk management, synthetic data simulation, and investigator- and data-science-driven prompting further strengthen resilience and operational efficiency.

According to Gartner, "By 2028, **60%** of software engineering teams will use AI evaluation and observability platforms to build user trust in AI applications, up from **18%** in 2025.[16]"

These capabilities must operate consistently across onboarding, authentication, account monitoring, transaction monitoring and credit decisioning — areas that are increasingly converging.

In an agentic economy, fraud prevention becomes the continuous governance of trust between humans, agents and institutions.

The new identity stack

Agent identity: AI agent provenance, permissions, behavioural norms

Ecosystem: digital identity wallets, open banking data

Dynamic: behavioural biometrics, transaction context

Foundational: documents, physical biometrics, device signals, fraud consortia matching

Read more Experian Global Insights

[16]*Gartner, Market Guide for AI Evaluation and Observability Platforms, Manjunath Bhat, Alex Coqueiro, et al., 2 February 2026*

Contributors:

Mihail Blagoev, *Lead Global Solution Strategy Analyst*
David Britton, *SVP, Strategy & Business Development*
Rebecca McGrath, *Global Content Marketing Manager*
Toby Sewell, *Global Product Marketing Manager*
Matthew Stennett, *Brand Design Manager*
Michael Touchton, *Senior Manager of Analyst Relations*