



# IMF

---

# NOTES

---

## **How Agentic AI Will Reshape Payments**

Prepared by Sonja Davidovic and Hervé Tourpe

©2026 International Monetary Fund

## How Agentic AI Will Reshape Payments

NOTE/2026/004

Sonja Davidovic and Hervé Tourpe

**DISCLAIMER:** The IMF Notes Series aims to quickly disseminate succinct IMF analysis on critical economic issues to member countries and the broader policy community. The views expressed in IMF Notes are those of the author(s), although they do not necessarily represent the views of the IMF, or its Executive Board, or its Management.

**RECOMMENDED CITATION:** Davidovic, Sonja, and Hervé Tourpe. 2026. "How Agentic AI Will Reshape Payments." IMF Note 2026/004, International Monetary Fund, Washington, DC.

Publication orders may be placed online, by fax, or through the mail:

International Monetary Fund, Publications Services  
P.O. Box 92780, Washington, DC 20090, USA  
Tel.: (202) 623-7430 Fax: (202) 623-7201  
Email: [publications@imf.org](mailto:publications@imf.org)  
[bookstore.IMF.org](http://bookstore.IMF.org)  
[elibrary.IMF.org](http://elibrary.IMF.org)

# Contents

- Executive Summary ..... 4
- Introduction ..... 5
- Background..... 5
  - Motivation and Scope..... 5
  - What Is Agentic AI?..... 6
  - Brief History of AI in Payments ..... 6
  - Table 1. Evolution of the Use of AI in Payment Systems ..... 7
- Trustworthy Agents ..... 8
  - AI -Payment Layered Model ..... 8
    - Layer 1—Intent and Orchestration Layer ..... 8
    - Layer 2—Control and Authorization Layer ..... 9
    - Layer 3—Settlement Layer ..... 10
  - Implications of Functional Separation ..... 10
  - Table 2. Mapping of the Payment Journey with the Three-Layer Model ..... 11
- Agentic AI Capabilities in Payments ..... 12
  - Evolution of E-Commerce ..... 12
  - Automated Cross-Border Payment Flows and Liquidity Management ..... 13
  - Agentic Compliance Solutions ..... 14
- Risks, Gaps, and Mitigation Strategies..... 15
  - Table 3. Risk Classification Matrix—Agentic AI in Payments ..... 17
  - Mitigation Strategies..... 18
    - Systemic Measures ..... 18
    - Private Sector Measures ..... 19
    - Public Sector Measures ..... 19
- Conclusion ..... 20
- References ..... 22

# How Agentic AI Will Reshape Payments

Prepared by Sonja Davidovic and Hervé Tourpe\*

April 2026

## Executive Summary

Artificial intelligence (AI) is entering a new phase in which systems can act autonomously on behalf of users. These “agentic” AI systems can interpret objectives, plan multistep actions, and interact with digital services with limited human intervention. In payments, this shift could move transaction initiation from explicitly human instructions toward agent-mediated decision making. Although adoption remains at an early stage, experimentation by technology firms, payment networks, and financial institutions shows that agentic models will become increasingly relevant over time.

This note takes stock of these developments and examines how agentic AI could affect the functioning of payment systems, including authorization, liquidity management, settlement, compliance, and operational resilience. It does not seek to draw definitive conclusions or propose prescriptive policy measures. Instead, it aims to frame key design questions, architectural tensions, and risk channels that may warrant attention as adoption evolves.

A central challenge is the interaction between probabilistic, adaptive decision making and the deterministic requirements of payment infrastructures. To structure the analysis, the note introduces a three-layer conceptual framework separating (1) intent formation and orchestration, (2) authorization and control, and (3) settlement. The framework is a normative analytical lens, informed by emerging practice, to clarify where agentic capabilities can operate productively and where rule-based safeguards remain essential.

The note reviews potential use cases for agentic AI in payments and highlights risks related to authorization traceability, opacity, correlated agent behavior, cybersecurity, and unresolved legal and liability questions. It also discusses emerging mitigation approaches, including mandate-based authorization, architectural separation of decision making and execution, agent identity frameworks, programmable payment controls, audit trails, and tiered human-in-the-loop models. The note argues that the impact of agentic AI on payments will depend not only on technology but also on institutional design and governance choices as experimentation continues.

---

\* The authors would like to thank the following colleagues for helpful research material, discussions, and comments: Anita Angelovska Bezhoska, Marianne Bechara, Alexander Copestake, Clement Couchevellou, Era Dabla-Norris, Jose Deodoro, Jose Garrido, Kathleen Kao, Yosuke Kido, Pearl Kuebel, Baoping Shang, and Alexandre Balduino Sollaci.

## Introduction

---

Artificial intelligence (AI) is rapidly evolving from a tool that assists financial decision making into a technology capable of acting on behalf of economic actors. A new generation of agentic AI systems, software agents able to interpret objectives, plan multistep actions, and autonomously interact with digital services, has begun to emerge across e-commerce and financial markets. In payments, these systems are increasingly able not only to recommend transactions but also to initiate, coordinate, and manage financial operations under delegated authority. This development potentially represents a structural shift in the architecture of financial systems. Historically, payment infrastructures have been designed based on instructions initiated by individuals and processed by deterministic systems. Payment rails, from card networks to real-time gross settlement (RTGS) systems, rely on predictable rules, legal certainty, and clear accountability structures to ensure trust and financial stability. Agentic AI introduces a new element into this framework: probabilistic decision-making systems capable of initiating financial actions at machine speed.

Recent industry developments show that this transition is already underway. Major technology and payment companies are experimenting with agent-mediated commerce and payment flows, while new technical standards such as the Universal Commerce Protocol (UCP), Agent Payments Protocol (AP2), Agent-to-Agent (A2A) communication frameworks, and the Model Context Protocol (MCP) are rapidly emerging to enable interoperability among autonomous agents and payment infrastructures. Market forecasts suggest that agent-mediated commerce could generate significant economic activity in the coming decade, potentially transforming how consumers, businesses, and financial institutions interact with payment systems.

These developments raise important policy, architectural, and risk-management questions. Payment systems must reconcile two fundamentally different design logics: the adaptive, probabilistic nature of agentic AI systems and the deterministic requirements of financial market infrastructures. Without appropriate safeguards, delegating payment initiation to autonomous agents could introduce new operational, legal, and systemic risks, including misaligned incentives, model errors, and highly correlated automated behaviors across markets.

Given the early and still limited adoption of agentic AI in payments, this note does not seek to draw definitive conclusions or propose prescriptive policy measures. Instead, it takes stock of emerging developments and experimentation, and introduces a conceptual framework to structure the discussion of how agentic capabilities could interact with payment systems over time. The note reviews potential use cases, highlights key design questions and emerging risks associated with greater automation of payment initiation, and outlines governance and architectural considerations that may warrant attention as adoption evolves.

## Background

---

### Motivation and Scope

The rapid emergence of agent-mediated payment use cases raises a set of design and policy questions that existing payment frameworks were not built to address. Rather than users explicitly initiating transactions, these settings rely on software agents operating under delegated mandates to anticipate payment needs, evaluate options, and coordinate execution across multiple instruments and rails. They may even be authorized, under certain conditions, to make payment decisions.

This evolution can be characterized as a shift from explicitly human-initiated transactions (“click-to-pay”) toward agent-mediated decision processes (“decide-to-pay”), in which execution increasingly occurs at machine speed and across multiple layers of the payment value chain, subject to predefined objectives, constraints, and governance arrangements.

A growing set of industry actors, including payment networks, technology platforms (for example, open-source software such as Ethereum), and AI model providers are in a race to experiment with these capabilities. Although current implementations remain largely focused on improving how people find and compare products, experimentation is now rapidly expanding across a broad landscape of payment-related use cases, from fraud detection and compliance monitoring to treasury optimization and cross-border payment orchestration, reflecting the widening scope of agentic AI pilots across the ecosystem (Boston Consulting Group 2025). The current velocity of innovation means developments that once required years can now materialize within months, indicating that more rapid and substantive change could be imminent.

In the context of this, a central architectural challenge emerges. Core payment infrastructures are built on deterministic logic, requiring predictability, auditability, and legal enforceability at each step of the transaction lifecycle. Agentic AI systems, by contrast, rely on probabilistic reasoning and adaptive decision making, which can yield different outcomes under similar conditions. This paper analyzes how and where these fundamentally different properties can be reconciled, and a three-layer model is proposed to assess the appropriate scope for integrating agentic AI into payment workflows.

### **What Is Agentic AI?**

Agentic AI refers to autonomous systems that sense their environment, set goals, and perform multistep tasks with little human input. For example, in the commercial sector, such systems may track household gas tank levels, identify the most cost-effective propane supplier, and automatically arrange for a refill. Within warehouse operations, these technologies can detect incoming deliveries, retrieve corresponding billing information, and initiate payment instructions in accordance with predefined protocols. Furthermore, the Bank for International Settlements (BIS 2025) notes that AI agents can independently manage liquidity and prioritize payments within real-time gross settlement systems, effectively mirroring established prudential cash management practices.

Unlike traditional AI models that provide static predictions (see Table 1) or require continuous human involvement, agentic AI systems integrate advanced capabilities such as planning, dynamic adaptation, and tool orchestration, thereby enabling these systems to operate as self-directed agents within complex ecosystems (FinRegLab 2025). Technology protocols and standards are evolving very rapidly to enable AI agents to interact with any payment actor and data source (Table 2).

### **Brief History of AI in Payments**

Early AI applications in payments emerged in the 1980s through expert systems that encoded human logic into static “if-then” rules (Table 1). One illustrative example is American Express’ Authorizer’s Assistant, which automated elements of credit authorization. Although effective in controlled settings, such systems proved ill-suited to keep up with malicious behaviors evolving faster than rules could be updated.

As transaction volumes grew in the 1990s, payment providers increasingly turned to statistical machine learning. Systems such as Hecht-Nielsen Neurocomputing Corporation Software’s Falcon Fraud Manager and Visa’s real-time authorization tools used neural networks to assess fraud risk at scale.<sup>1</sup> These approaches introduced probabilistic reasoning into payment workflows, but they remained confined to detection and scoring rather than acting to remediate the issue.

---

<sup>1</sup> Neural networks are a class of machine-learning models inspired by biological neurons, which learn statistical patterns from large data sets through weighted connections rather than explicit, human-defined rules. In the 1990s, they were among the first techniques capable of adapting fraud detection models dynamically as transaction patterns evolved.

Subsequent waves, including graph-based analytics in the 2000s and deep learning in the 2010s, further enhanced fraud detection, authentication, and risk management, supporting innovations such as biometric payments and tokenization.<sup>2</sup> Across these phases, AI materially improved how payment decisions were evaluated, but payment initiation and settlement remained explicitly human driven.

**Table 1. Evolution of the Use of AI in Payment Systems**

Period	Key Technology	Innovations (Examples)	Function in Payment
1980s	Expert systems	American Express' Authorizer's Assistant	Rules-based credit authorization and fraud detection, using simple "if-then" logic.
1990s	Machine learning, early neural networks	Falcon Fraud Manager, convolutional neural networks for check processing	Probabilistic fraud scoring and pattern recognition; transaction-level risk assessment.
2000s	Graph analysis, anomaly detection	PayPal Igor and Ilya	Network based fraud and risk analysis through relationship mapping across accounts and devices.
2010s	Deep learning, biometrics	Apple Pay, Stripe radar	Biometric authentication, tokenization, advanced fraud detection, AI-assisted KYC, and identity verification.
2020s	Foundation models, conversational AI	"Hello UPI" (voice based UPI payments, India)	AI-mediated payment initiation, customer interaction, and workflow automation; execution remains deterministic.

Source: Authors.

Note: AI = artificial intelligence; KYC = know your customer; UPI = Unified Payments Interface.

The new generation of AI systems, called Large Language Models (LLMs), powers AI agents which are inherently nondeterministic: rather than producing a single, reproducible output for a given input, they generate responses by sampling from probability distributions over possible next tokens.<sup>3</sup> As a result, identical prompts can yield different outputs. This property has fundamental implications for where and how agentic AI can be trusted within the payment chain, particularly in relation to authorization, execution, and settlement functions.<sup>4</sup>

Nondeterminism also manifests through hallucinations, where models generate plausible but incorrect statements. Although successive model generations show reduced hallucination rates, the risk has not been eliminated and remains a key concern in payment, compliance, and settlement contexts (Omar and others 2025).

<sup>2</sup> Deep learning refers to a class of machine-learning techniques based on multilayer (or "deep") neural networks that automatically learn hierarchical representations from large volumes of data. In the 2010s, advances in computing power, data availability, and training methods enabled these models to outperform earlier approaches in tasks such as image, speech, and behavioral pattern recognition.

<sup>3</sup> In the context of AI, a "token" refers to a unit of text processed by a model and should not be confused with "tokenization" in the context of distributed ledger technology (DLT), where the term denotes the representation of assets, rights, or value on a ledger.

<sup>4</sup> One key control is "temperature," an AI setting that governs output variability. Low temperatures bias the model toward the most likely tokens, increasing predictability; higher temperatures allow greater exploration and diversity. In payment and financial contexts, where accuracy, auditability, and legal certainty are paramount, temperatures are typically set very low.

## Trustworthy Agents

---

AI agents are therefore very powerful, but constrained by their probabilistic nature, restricting their use in certain tasks in the payment chain. By contrast, payment infrastructures such as RTGS systems, card networks, instant payment platforms, and distributed ledgers operate deterministically:<sup>5</sup> transactions follow preset rules, outcomes are binary, and legal finality is guaranteed.<sup>6</sup>

This difference creates a structural challenge. The main risk with agentic payments does not come from using probability-based reasoning itself, but from letting adaptive systems make irreversible payments without proper controls, checks, or accountability. Therefore, the central issue is not whether AI should be used in payments - as AI has already been used for more than four decades (see Table 1) - but how to keep uncertain, probabilistic decision making separate from automatic payment execution. The next section introduces a three layer-model to clarify these roles and architectures.

### AI-Payment Layered Model

Traditional payment systems tie authorization directly to explicit instructions from a user or institution, who starts a payment that is then checked against established rules. By contrast, agentic payment models rely on objectives and constraints to express intent. For instance, a buyer might instruct an agent to “buy the latest book on topic x when it becomes available on any platform, at the best price.” Here, the agent decides when and where to search for the book and compares prices from various merchants before suggesting optimal choices. Despite these advances, humans typically remain involved in reviewing and finalizing purchases, as the mix of probabilistic decision making and deterministic execution can complicate accountability if issues arise.

To maintain clear accountability, one approach involves creating a defined boundary between AI-driven decisions and actual payment execution. This means separating the unpredictable aspects of decision making from the predictable steps of carrying out payments by organizing each into distinct layers. The three-layer model is intended as a normative conceptual framework, informed by emerging industry practice. It is not meant to describe all existing payment architectures, but to highlight a design principle: concentrating probabilistic, adaptive reasoning upstream, while preserving deterministic authorization and settlement where legal finality and systemic stability are required. We also use the model to highlight how new technology and standards are evolving to support each layer.

### Layer 1—Intent and Orchestration Layer

This layer includes the probabilistic agentic systems and protocols that translate high-level user objectives, or “intent”, into structured, machine-readable instructions. Technologies at this layer enable reasoning, planning, search, negotiation, and multi-agent coordination capabilities, without performing any authorization or execution. Table 2 shows how new standards are rapidly emerging, and are being adopted, extending the capabilities of LLM-based agents. For example, the MCP standardizes agents’ access to external data and tools, while A2A protocols enable interoperability and coordination among agents developed by different vendors. The x402 standard builds on the HTTP 402 web protocol and allows agents to embed payment requirements directly within HTTP requests. This allows to automatically negotiate and handle paid services over the internet. One of

---

<sup>5</sup> In this framework, “deterministic” refers to the execution of authorization and control rules, but not to how those rules are defined. Regulatory thresholds, sanctions lists, and smart-contract constraints reflect human judgment and policy choices, but once established, they are enforced through fixed, reproducible decision procedures. Therefore, human discretion remains essential at the level of rule design and governance, while deterministic execution underpins auditability, accountability, and stability.

<sup>6</sup> Although AI models can be configured to behave more predictably, this does not make their decision making rule-bound in the sense required for authorization or settlement. Therefore, the concern is not variability alone but also the absence of guaranteed reproducibility and legal interpretability at the level of individual transactions.

the most impactful standards is Google's UCP (Google 2026), which provides a shared grammar for discovery, comparison, and the ability for agents to create and manage post-purchase logic.

Industry pilots already leverage these new standards. For example, Visa's Intelligent Commerce and Mastercard's Agent Pay test agent-initiated shopping and payment flows, where agents autonomously construct purchase intent under predefined limits. Thus, Layer 1 serves as the locus for adaptive orchestration and delegated initiation, producing structured intent that must still pass through the deterministic authorization and settlement layers, which will be described next.

## **Layer 2—Control and Authorization Layer**

This layer enforces deterministic constraints that govern whether actions proposed or initiated by agents may proceed toward execution. Technologies in this layer ensure that authorization decisions are ultimately governed by deterministic policy rules, even when informed by upstream probabilistic systems. Research in this domain focuses on protocols that shift trust from human oversight to technical safeguards through verifiable claims, authorization constraints, and identity frameworks (Hu and Rong 2025).

The core mechanism at this layer is AP2, which binds agent-initiated actions to cryptographically verifiable mandates specifying scope, limits, actor identity, and permitted conditions. The AP2 standard also supports extension x402 which enables stablecoin integration. These mandates ensure that downstream authorization reflects explicit user consent rather than model-generated inference. Some more established web authorization technologies, such as OAuth 2.0 or OpenID Connect, are often used to enable robust Know Your Agent verification through attested agent identities (World Economic Forum 2026).

Recent industry implementations illustrate how mandate based authorization can be operationalized in agent-initiated payment flows. For example, some payment service providers have introduced tokenized authorization mechanisms that allow AI agents to initiate transactions using a user's preapproved payment methods (including card and non-card options) without accessing underlying credentials (Stripe 2026). Such approaches demonstrate how structural authorization and payment method choice can be preserved while maintaining deterministic controls over execution.

Efforts to connect agentic payments with distributed ledger technology (DLT)-based authorization and controls are also advancing along two complementary tracks. Verifiable claim mechanisms, such as ERC-1812, enable off chain, securely signed attestations that can be verified by smart contracts or policy engines to support deterministic checks on identity attributes, entitlements, or constraints before a payment is authorized. In parallel, programmable wallet and account-abstraction standards such as ERC-6900 enforce spend limits, velocity controls, counterparty restrictions, and approval workflows directly at the wallet layer, creating a deterministic control safety gate between intent formation and execution. Emerging proposals such as ERC-8004 extend this authorization framework to the agents themselves by defining on-chain registries for agent identity, validation, and reputation, supporting decisions about which agents are permitted to act, and strengthening trust in machine-initiated actions within deterministic payment infrastructures.

Traditional network-level controls, such as issuer rules, anti-money laundering (AML)/combating the financing of terrorism filters, velocity checks, tokenization rules, sanctions screening, and dispute guardrails, operate in parallel, providing the deterministic validations. Supervisory intelligence systems may inform these controls, such as AML detection models (FinRegLab 2025), but they do not perform authorization themselves.

However, most payment regimes require that a payment order be traceable to an authorized instruction from an account holder or its legally recognized agent. Agent-initiated payments challenge this model, as individual transactions may not correspond to explicit, transaction-level instructions. Authorization instead becomes structural and mandate based, which raises questions about traceability, consent, and liability under existing

legal frameworks. This shows that broader and legally workable concepts of authorization (grounded in verifiable mandates, scope limitations, and auditability) are needed as agentic payment models evolve.

Taken together, these technologies define Layer 2 as a strictly rules-based authorization boundary: it accepts structured intent from Layer 1 only if it satisfies verifiable mandates, policy constraints, and regulatory checks. Anything accepted at this layer becomes an authorized payment instruction eligible for deterministic execution in Layer 3. Anything that fails is rejected or routed back to the agent for revision. This separation ensures accountability, auditability, and compliance even as agentic systems introduce greater automation upstream.

### **Layer 3—Settlement Layer**

This layer includes traditional deterministic settlement infrastructures, such as RTGS systems, instant payment networks, card network clearing engines, and newer settlement systems such as central bank digital currency platforms and distributed ledger-based settlement rails. Layer 3 executes payment instructions with irrevocable legal finality through settlement infrastructures. Where Layer 3 corresponds to a financial market infrastructure, its operation is subject to the BIS CPMI-IOSCO Principles for Financial Market Infrastructures, which set standards for governance, risk management, access, and settlement finality. In contrast to the adaptive logic of Layers 1 and 2, technologies in this layer are explicitly designed for predictable, rules-bound execution, operational resilience, and strict auditability, consistent with the principles used by financial market infrastructures.

Execution on this layer also relies on settlement-native technologies such as programmable wallets and token standards. In DLT environments, such as those used by stablecoin or central bank digital currencies, ERC 4337, launched in 2023, provides smart contract wallet execution for authorized transactions. These constraints remain rule based: they enforce policies defined in Layer 2 but do not interpret user objectives or run adaptive reasoning.

In this architecture, Layer 3 is the final, non-probabilistic endpoint of the payment chain: it takes only those instructions that have passed deterministic controls in Layer 2 and executes them without modification, optimization, or reinterpretation. Typically, agentic algorithms do not operate here.<sup>7</sup> This preserves legal certainty, limits systemic risk, and ensures that the foundations of the payment system remain stable, synchronized, and trustworthy even as upstream processes become increasingly automated.

### **Implications of Functional Separation**

The functional separation introduced earlier has several analytical implications for how agentic AI interacts with payment systems. First, it explains why near-term value from agentic AI is concentrated upstream of settlement, where probabilistic reasoning supports optimization and orchestration. Second, it clarifies why settlement systems are unlikely to become “intelligent” in the same sense as agents. Although they may support conditional execution (for example, with smart contracts on distributed ledgers), their core function in that layer remains deterministic finality. Third, it highlights a shift in risk from individual transaction correctness toward system-level behavior.

---

<sup>7</sup> Although limited agentic functions, such as automated validation or reconciliation, can in principle operate at the settlement layer, doing so introduces trade-offs between efficiency and legal certainty. Therefore, the framework treats Layer 3 as non-probabilistic by design, not because agentic execution is impossible, but because constraining it helps preserve accountability and finality at system scale.

**Table 2. Mapping of the Payment Journey with the Three-Layer Model***(Layers: L1 = Intent and Orchestration, L2 = Control and Authorization, L3 = Settlement)*

<b>Payment Channel Function</b>	<b>Problem Addressed</b>	<b>Technologies, Standards, Protocols</b>	<b>Layer(s)</b>
<b>Intent and context management</b>	Enable agents to reason over user objectives, constraints, preferences, and state across multistep tasks	LLMs; MCP for contextual tool access; ACP connects buyer, their agent, and business	L1
<b>Agent coordination</b>	Allow multiple agents (buyer, merchant, treasury, compliance, risk) to exchange plans, negotiate actions, and delegate subtasks	A2A Protocol; multi-agent orchestration frameworks	L1
<b>Commerce workflow orchestration</b>	Standardize discovery → comparison → offer → prepares for authorization	Universal Commerce Protocol	L1 + L2
<b>Delegation and authorization</b>	Prove that an agent is authorized to act for a user/institution with scope-limited permissions, revocation, and auditability	AP2 (digitally signed mandates and proof of intent); OAuth 2.0; OpenID Connect	L2
<b>Agent identity (know your agent)</b>	Identify and authenticate software agents as distinct operational actors	ERC 780 claims registry; ERC-8004 for identity and reputation registries; AP2 verifiable agent credentials	L2
<b>Real-time compliance and fraud risk filtering</b>	Detect anomalies, sanctions risk, or policy violations before authorization	ML/AI fraud systems; mandate constraints through AP2; adaptive AML models	L2
<b>Programmable settlement controls</b>	Encode agent permissions, spending limits, guardrails, and conditions directly into digital money or wallets	ERC 4337 smart contract wallets; ERC 6900 modular smart accounts; ERC 1812 off-chain authorizations	L2
<b>Platform-specific agent execution</b>	Allow agents to initiate payments inside closed or semi-closed ecosystems	AP2 Protocol; x402 on-chain payment request	L2 + L3
<b>Programmable money and digital asset rails</b>	Support agent-driven payments with programmable conditions and settlement guarantees	DLT; smart contract layers; regulated token frameworks	L3
<b>Traditional payment rails</b>	Plug into traditional, regulated payment systems	Platforms for RTGS, credit cards, and so on	L3
<b>Liquidity management</b>	Optimization of liquidity management	Platforms for RTGS, stablecoins	L3

Source: Authors.

Note: A2A = Agent-to-Agent; ACP = Agentic Commerce Protocol; AI = artificial intelligence; AML = anti-money laundering; AP2 = Agent Payments Protocol; DLT = distributed ledger technology; LLMs = Large Language Models; MCP = Model Context Protocol; ML = machine learning; RTGS = real-time gross settlement.

## Agentic AI Capabilities in Payments

---

Agentic AI in payments can introduce benefits across a wide area of use cases. Experts describe agentic AI as “digital factories” of AI agents that handle entire tasks, with humans only needed for exceptions and oversight (McKinsey & Company 2025b). The autonomous management and execution of complex tasks for payment-related use cases can lower transaction and operational costs, improve liquidity, enhance compliance processes, and reduce fraud. Removing human latency and administrative friction is expected to enhance efficiency, accelerate capital circulation, and boost productivity.

### Evolution of E-Commerce

AI agents are changing e-commerce by replacing simple automation with autonomous reasoning. With reduced human involvement, transactions can accelerate and consumption trends may shift. Agentic AI could enhance financial inclusion by reducing search costs and information asymmetries, allowing users to compare financial products, fees, and terms more effectively (Cao 2026). As agents prioritize low-latency payment solutions and remove inefficiencies, they could directly increase money velocity.

Several companies have launched agentic commerce initiatives to test potential efficiency gains for both operations and consumers (Box 1). In e-commerce, agents can reduce transactional friction by autonomously managing the purchase process, from product search and price comparison to discount application, availability verification, and execution. For merchants, this automation can accelerate purchase cycles, with evidence suggesting that LLM-based agents significantly shorten consumer interaction times in digital marketplaces (Yan and others 2025). Consumers benefit from personalized decision support that incorporates preferences, constraints, and real-time price signals. Agents can leverage the richer context provided in the user’s prompt to anticipate demand patterns and transmit structured feedback to merchants (Beard 2025). Besides lowering cognitive burdens for users, this personalization can increase customer lifetime value through more relevant product matching and streamlined purchasing experiences.

Agentic systems may also extend beyond checkout to post-purchase functions, including delivery coordination and returns management. Gartner (2025) estimates that AI agents can autonomously resolve up to 80 percent of common customer service issues by 2029. For merchants, automated returns processing and adaptive pricing mechanisms can reduce service costs, improve dispute resolution, and enable faster responses to shifts in demand and inventory conditions (McKinsey 2025b). Reflecting these expected gains, surveys indicate that firms are prioritizing AI investments in customer service, marketing, and product development (Microsoft 2025).

More broadly, agentic AI commerce services can spur modernization of e-commerce platforms as they enable integration across AI systems. For instance, PayPal recently announced strategic partnerships allowing merchants to seamlessly enable product discovery on AI platforms (PayPal 2025). There could also be important workforce consideration as agents could replace or augment operational roles (including merchandising, supply chain management, and product optimization), thus reducing operating costs and boosting productivity (Brynjolfsson, Li, and Raymond 2023).

## **BOX 1. Illustrative Market Experimentation with Agentic Payment Workflows**

Recent experimentation by technology firms and payment networks provides early evidence of how agentic AI capabilities are being integrated into payment initiation and orchestration layers. These initiatives offer insight into emerging design patterns rather than settled or standardized architectures.

### **OpenAI/Stripe (Agentic Commerce Protocol)**

OpenAI's "Instant Checkout" in ChatGPT, powered by the Agentic Commerce Protocol, enables users to buy products directly within a conversation. Although it uses standard merchant acquiring infrastructure, it introduced a new economic model in early 2026, charging a 4 percent transaction fee for autonomous agent-led conversions (Stripe 2025).

### **Amazon (Rufus and "Buy for Me")**

Amazon has transitioned from simple recommendations to delegated purchasing. Its "Buy for Me" feature enables the Rufus assistant to navigate external websites and complete transactions on behalf of the customer, positioning the shopping agent as a primary interaction surface rather than a support tool.

### **Google (Universal Commerce Protocol)**

Launched in January 2026, Universal Commerce Protocol standardizes how businesses connect with AI agents across the shopping journey. It enables "Native Checkout" within Google Search's AI Mode and Gemini, allowing users to purchase from retailers like Etsy and Wayfair without leaving the AI surface.

### **Visa and Mastercard (Network-Level Control)**

Visa's "Intelligent Commerce" and Mastercard's "Agent Suite" (launched in the second quarter of 2026) focus on "Know Your Agent" frameworks. These initiatives provide the registration, cryptographic signatures, and network tokens required to distinguish legitimate agents from malicious bots, ensuring that deterministic authorization stays within established rails.

### **PayPal (Cymbio Infrastructure)**

Through its 2026 acquisition of Cymbio, PayPal has positioned itself as the "trust layer" for the agentic web. This allows independent agents to use PayPal's transaction graph and secure vaults to facilitate settlement while maintaining merchant-of-record status for the retailer.

## **Automated Cross-Border Payment Flows and Liquidity Management**

Agentic AI systems go beyond smart automation. They can coordinate and execute multistep workflows across financial distributed networks autonomously. Agentic AI can orchestrate the entire cross-border payment chain from payment initiation, optimizing routing options (including correspondent banks, local partners, and tokenized rails), triggering compliance checks, and monitoring settlement and post-settlement exceptions, as observed in e-commerce use cases. Proponents believe that this automated payment flow can reduce delays associated with manual intervention and rigid workflows (Convera 2025). The argument is convincing: the analysis of real-time transaction data and the ability to reason over past experiences and changing context (including varying cost, fees, and channel performance) provide AI agents with the ability to dynamically select the most efficient payment path and to adjust to specific situations (Capco 2025). Agentic AI can also play an active role in

autonomously managing liquidity when executing transactions, such as reallocating funds based on predefined parameters and real-time market conditions. BIS's findings show that generative AI systems can fulfill cash management functions such as maintaining precautionary liquidity buffers, prioritizing urgent payments, and balancing trade-offs between liquidity costs and settlement delays without specialized training (BIS 2025).

Similarly, AI agents can help streamline foreign exchange (FX) management as one of the core functions in cross-border payments. With the help of agentic AI, financial institutions and companies can continuously monitor real-time exchange rates, analyze spreads across banking rails, optimize timing for conversion, and choose cost-effective paths for transferring funds across multiple currencies (Uppuduri 2025). LLMs can enhance predictive analytics by extracting signals from unstructured data sources that traditional models may underuse, supporting improved forecasting and risk-management functions in areas such as liquidity planning and foreign-exchange management. These capabilities could generate cost savings and operational efficiencies, as piloted by Citi and Ant International in an AI-powered tool aimed at reducing FX-hedging costs (Reuters 2025). There are also potential agentic AI applications for off-balance sheet transactions, such as FX derivatives, which are out of scope for this note.

The integration of different payment systems and networks is particularly important for cross-border payments given the jurisdiction-specific payment system landscape. Even though network initiatives such as SWIFT Global Payments Innovation and Visa B2B Connect are not agentic AI, they set the stage for agents to leverage various rails across jurisdictions (Scalefocus 2025). As stablecoins are becoming more prevalent in cross-border payments, AI agents using stablecoin rails might become an important infrastructure leveraging the synergies of DLT functionalities and agentic AI as evidenced by standard protocols such as AP2 enabling stablecoin integration (Desai 2025). The convergence of AI and stablecoins has the potential to power a more inclusive and efficient "internet financial system" that accelerates the evolution of global finance (WEF 2025).

### Agentic Compliance Solutions

Agentic AI has the potential to significantly improve compliance processes by embedding regulatory logic directly into operational workflows. Unlike traditional automation tools, agentic systems can interpret objectives, monitor activity in real time, and autonomously take actions within predefined guardrails such as flagging suspicious transactions, escalating high-risk cases, or adjusting controls when regulatory thresholds are met. By integrating continuous monitoring, explainability, and audit trails into their core architecture, agentic AI enables compliance functions to operate at the same speed and scale as modern digital systems, reducing operational burden and human error while strengthening consistency, traceability, and regulatory alignment.

Agents allow for real-time compliance monitoring that continuously evaluates transactions and activities instead of relying on periodic or batch reviews. Several companies have introduced AI-based compliance monitoring and fraud detection tools. Mastercard's AI-powered "Decision Intelligence" system evaluates transactions in milliseconds for fraud risk allowing for real-time compliance screening at transaction speed assessing contextual risk before authorization (Mastercard 2024). Visa developed a real-time risk-scoring solution for account-to-account payments that scores transactions in milliseconds based on real-time contextual data to automatically approve, decline, or flag transactions (Visa 2026). It embeds compliance logic such as fraud, sanctions flags, and risk rules into real-time transaction flows, thereby introducing compliance by design at network scale.

Agentic AI systems introduce the capability of automated rule enforcement applying regulatory requirements and governance frameworks programmatically at the point of decision. An academic prototype presents a reference architecture for agentic AI handling AML and Know-Your-Customer (KYC) compliance tasks with integrated explainability and traceability. It shows how agentic compliance can embed, interpret, and enforce regulatory logic directly into autonomous systems without manual intervention (Axelsen, Licht, and Damsgaard 2025). By embedding identity safeguards such as provenance verification and policy-based access control

directly into the code, institutions can turn regulation from a constraint into a practical system level enabler (Preis 2026).

AI governance by design embeds ethical, legal, and societal values into agentic systems, enabling proactive risk management (Joshi 2025). These systems help automate compliance tasks, cut errors, reduce costs, and scale operations with autonomous agents handling routine work while escalating complex cases to humans (McKinsey & Company 2025a). AI for fraud detection can become an element of a digital financial infrastructure as is the case in India's Unified Payment Interface. Unified Payment Interface integrates payments, data, and intelligence as a unified stack, with embedded AI for real-time fraud detection, alternative credit scoring, and automated reconciliation. This architecture supports domestic and international expansion through Digital Public Infrastructure exports (Shivam 2026).

Multi-agent systems can distribute compliance tasks across specialized domains, whereas federated coordination ensures coherence across jurisdictions (Onyekaonwu, Igba, and PeterAnyebe 2024). Agents communicate through secure protocols, maintaining data sovereignty while exchanging regulatory intelligence. Applying this to compliance, one specialized agent can perform regulatory change scanning, another risk scoring, a third control mapping, and a fourth execution of remediation, all orchestrated by a coordinator that handles task delegation, conflict resolution, fallback, and prioritization. Based on the proposed three-layer framework, one agent could initiate a compliance check in the intent/orchestration layer, another could scan and flag any changes in regulatory requirements in the control layer, while a third agent could issue a validation of a compliance check in the settlement layer.

## Risks, Gaps, and Mitigation Strategies

---

Agentic AI systems are capable of autonomously interpreting objectives, taking multi-step actions, and adapting dynamically to changing environments. They represent the next wave of innovation in financial services. In consumer finance, payments, compliance, and supervision, they could significantly reduce cognitive and operational burdens by automating complex financial and supervisory tasks. However, their autonomy, opacity, and non-deterministic behavior introduce material risks to consumer protection, market stability, and regulatory oversight. Although agentic AI has broad implications across economic activity, the focus lies on how these technologies affect the functioning of payment systems, including liquidity management, authorization, settlement, and the operational stability of payment infrastructures. Table 3 classifies the main risks according to their source and potential impact.

The convenience of agentic AI for e-commerce introduces risks to consumer autonomy. Agentic systems may misinterpret user intent, optimize provider incentives rather than user welfare, drift from original objectives over time, or engage in subtle behavioral nudging at scale. Communication challenges between agents and end users might also arise because consumers do not always have a clear sense of their own short- and long-term financial goals or risk appetites (Aldasoro and others 2024). In multi-agent environments, there could be a misalignment in objectives between different elements or layers of an agentic system, or divergence between individual agent goals (Carichon 2025). Algorithmic herding occurs if dominant models identify identical market signals, while simultaneous execution could trigger flash crashes that bypass traditional circuit breakers (Ogbuonyalu and others 2024). Algorithmic herding can impair the functioning of payment systems by synchronizing liquidity demand, amplifying procyclical behavior, and creating congestion across payment rails, thereby undermining the predictability and resilience of core financial market infrastructures (BIS 2024). Too many economic agents being able to autonomously trigger and process financial operations creates a significant risk if malicious actors can influence or control agents through prompt injection attacks (Hogan Lovells 2025).

Ongoing access to financial data, ability to engage in behavioral nudging, and capacity for autonomous action can potentially increase the scale, subtlety, and personalization of activities that run counter to consumers'

interest (FinRegLab 2025). These dynamics have already arisen in the context of some AI-driven personal finance apps, which have been criticized for recommending cash advance offers and promoting borrowing products when a user's account balance dips below particular thresholds, even in situations where those products would worsen consumers' long-term financial outcomes (Aldasoro and others 2024).

Generative AI models are prone to "hallucinations," creating false information with high confidence. In financial contexts, such errors can have significant consequences (Aldasoro and others 2024). The black-box nature of many AI systems further complicates accountability and redress. The opacity and nonlinearity of AI systems make it difficult for supervisors to fully understand model outputs, limiting their ability to detect potential systemic risks in a timely manner. High correlation in agents' behavior could also pose significant systemic risks (Aldasoro and others 2024). AI might enable more sophisticated risk assessment models and improve the prediction of institutional failures, spot market manipulation, or identify compliance-related issues. Although generative AI can be powerful for regulatory reporting and compliance through interactive communication with agents, the routine use of such methods is not established.

Another significant risk relates to data security and privacy protection. Autonomous agents relying on third-party services such as cloud providers, AI model endpoints, and financial services require user's sensitive data such as bank credentials, credit card numbers, and crypto wallet keys, exposing the user to data leaks and privacy concerns and creating a highly concentrated point of vulnerability (Aldasoro and others 2024).

There are also concerns about market concentration and competition. Generative AI is fed by vast amounts of data which require computing power, which can only be provided by a few, dominant companies, thereby introducing concentration risks (Korinek and Vipra 2025). Most of the AI supply chains show a high degree of concentration, from data centers to cloud computing and AI applications. Such concentration can threaten innovation and raise financial stability, operational, and reputational risks (BIS 2025). Moreover, as algorithms become more standardized and are uniformly used, the risk of herding behavior and procyclicality grows (Bank of England 2022). Model herding relates to the inadvertent use of similar optimization algorithms which can contribute to flash crashes, increase market volatility, and contribute to illiquidity during times of stress (Aldasoro and others 2024). Similar to algorithmic herding observed in financial markets, correlated behavior in payment initiation or liquidity optimization can lead to synchronized payment flows, thus increasing intraday liquidity demand and potentially straining settlement capacity and system efficiency (Kabadjova and others 2023). Finally, there are challenges because of increased competition in the banking sector as agents are continuously sweeping funds to high-yield accounts, eliminating the "inertia dividend," and forcing banks to compete on value instantly (McKinsey & Company 2025c).

Although the benefits of AI in cross-border payments are clear, financial institutions face significant challenges in integrating these technologies into their existing systems. Legacy infrastructure in many banks and financial institutions may not be compatible with cutting edge AI technologies, necessitating substantial investments in system upgrades and data migration (Scalefocus 2025). If such agents are used to automate high-speed FX execution or liquidity timing at scale, the existing evidence from high-frequency trading research indicates that more speculative high-frequency activity can worsen liquidity and raise intraday volatility, amplifying FX market swings (IMF 2025). Cross-border payments can benefit from synergies of DLT-based stablecoins and agentic AI, but there are associated risks that digital assets could amplify. Converting funds into cryptocurrencies to accelerate settlement can introduce additional volatility and custodial risks relative to traditional payment rails, as crypto assets typically exhibit higher price volatility and rely on specialized custody infrastructures outside the regulated banking system (Financial Stability Board 2022; IMF 2025). Crypto assets lack central trusted intermediaries and depend on decentralized consensus mechanisms, which alters the risk profile of asset transfer and exposes participants to price fluctuations and counterparty vulnerabilities. Even stablecoins have historically experienced peg instability and reserve-related counterparty risk under stress, underscoring that faster transfer speed comes with trade-offs in stability and custody (FSB 2023).

Besides risks, there are gaps that hinder the adoption of agentic AI in payments. There is a growing skills gap in the financial sector, with a shortage of professionals who possess both financial expertise and advanced AI knowledge (Uppuduri 2025). Many larger financial institutions have only recently begun to build internal expertise needed to safely integrate agentic AI into consumer-facing products (FinRegLab 2025).

More specific gaps relate to the authentication or KYC requirements for agents. Autonomous AI agents capable of initiating transactions challenge existing identity and authentication frameworks in payments. Traditional authorization mechanisms, including KYC processes and multifactor authentication, are designed around human users who explicitly approve transactions. When payments are initiated autonomously by software agents acting under delegated authority, verifying both the identity of the agent and the intent of the underlying user becomes significantly more complex. This shift raises new questions around authentication, accountability, and compliance in payment systems designed for human actors (Acharya 2025). There is also ambiguity regarding legal liability for autonomous agents.<sup>8</sup> Current liability regimes assume human intent and direct causation, but these become ambiguous when autonomous agents make decisions independently. When an agent acts on behalf of a principal in a transaction, determining liability in the event of disputes depends on whether the agent acted within its authority and whether the harm resulted from the agent’s own conduct or from the principal’s instructions. Therefore, both principals and agents may face liability depending on the circumstances, although the delineation of legal liability is unclear (Shukanayev 2025). There is a general lack of regulatory clarity on accountability and liability of agents in payments. Existing regulations struggle to distinguish between “unauthorized use” and “user negligence” if an agent hallucinates and misdirects funds (Aldasoro and others 2024).

**Table 3. Risk Classification Matrix—Agentic AI in Payments**

<b>Risk</b>	<b>(1) Primary Source of the Risk (Whose Actions)</b>	<b>(2) Who Bears the Cost If It Materializes</b>	<b>(3) Market Failure Justifying Policy Intervention?</b>
<b>Instruction gap (structural versus transactional authorization)</b>	Account holders delegating broad mandates; AI agents executing payments without transaction level instructions	Account holders (unexpected payments); PSPs (disputes); payment systems (operational strain)	Yes. Information and control asymmetry; misalignment between legal authorization models and agentic execution
<b>Opacity of agent decision making</b>	AI developers and deployers designing non-interpretable models; platforms integrating agents into payment flows	Users (lack of redress); PSPs (compliance failures); supervisors (reduced oversight capacity)	Yes. Information asymmetry and unverifiable decision processes
<b>Highspeed, machine time payment execution</b>	AI agents and platforms optimizing execution speed; inadequate throttling by PSPs	PSPs and payment systems (operational overload); end users (error propagation)	Yes. Externalities from speed amplification and coordination effects
<b>Authorization traceability failures</b>	PSPs and platforms relying on mandate based authorization without robust auditability	PSPs (liability exposure); users (disputed payments); courts/regulators (legal uncertainty)	Yes. Legal infrastructure mismatch; incomplete contracting
<b>Ambiguous liability allocation (agency based)</b>	Account holders, PSPs, and system operators interacting under outdated liability assumptions	PSPs and users (litigation, losses); payment systems (reputational risk)	Yes. Legal uncertainty and incomplete allocation of responsibility
<b>Product liability exposure from autonomous behavior</b>	AI model providers, integrators, and platforms deploying adaptive systems	PSPs, platforms, or users depending on legal interpretation; potentially model providers	Yes. Existing product liability regimes not designed for adaptive, post deployment behavior
<b>Correlated agent behavior (herding in payment timing or routing)</b>	Homogeneous models and shared optimization objectives across agents	Payment systems (liquidity stress); PSPs (settlement delays); end users (failed payments)	Yes. Coordination externalities and systemic risk

<sup>8</sup> In addition to agency-based liability, agentic AI systems may also raise issues under product liability regimes, particularly where autonomous behavior leads to harmful outcomes after deployment. Existing product liability frameworks are not yet well adapted to systems whose behavior may evolve over time, adding uncertainty to accountability and redress.

Risk	(1) Primary Source of the Risk (Whose Actions)	(2) Who Bears the Cost If It Materializes	(3) Market Failure Justifying Policy Intervention?
<b>Intraday liquidity stress and settlement congestion</b>	Agents optimizing payment timing and liquidity use; PSPs failing to impose constraints	Payment systems; central banks (liquidity backstops); participants	Yes. Systemic liquidity externalities
<b>Cybersecurity and attack surface expansion</b>	Platforms integrating agents with multiple tools/APIs; weak access controls	Users (fraud); PSPs (losses); payment systems (operational disruptions)	Yes. Security externalities and under-investment in resilience
<b>DLT settlement without legal finality</b>	System designers deploying DLT without statutory settlement recognition	Users and intermediaries (reversals, insolvency risk); legal system	Yes. Legal infrastructure gap and jurisdictional fragmentation
<b>Regulatory blind spots (KYA, supervision, monitoring)</b>	Regulators lagging technological change; platforms operating across borders	Financial system at large (loss of trust); regulators (oversight failure)	Yes. Public good nature of supervision and cross-border coordination problems

Source: Authors.

Note: AI = artificial intelligence; DLT = distributed ledger technology; KYA = know your agent; PSPs = payment service providers.

## Mitigation Strategies

Agentic AI systems are non-deterministic, meaning that identical inputs can lead to different outputs because of their probabilistic nature (NIST 2023). This characteristic introduces operational and financial risks when such systems initiate transactions or execute payment instructions. A key mitigation approach is the implementation of formal AI governance structures that define accountability, risk controls, and lifecycle monitoring for AI systems. The framework should emphasize transparency, accountability, and continuous monitoring throughout the AI lifecycle. The governance framework could include AI risk committees within financial institutions, board-level oversight of AI deployment, and internal model risk management adapted for agentic systems (NIST 2023).

## Systemic Measures

As autonomous agents can execute transactions at high speed, human approval or supervisory intervention should be required for high-risk or high-value transactions executed by AI agents. Research on agent governance frameworks emphasizes human-in-the-loop supervision as a core safeguard to monitor and intervene in autonomous agent decisions to reduce the risk of erroneous or malicious actions (Chiris and Mishra 2025).<sup>9</sup> This type of oversight would suggest transaction thresholds requiring human approval, supervisory dashboards for agent decisions, and manual override mechanisms. Once transactions reach legal finality, errors or unintended behavior may be difficult or impossible to reverse. This creates a strong case for ex-ante interruption and containment mechanisms, commonly referred to as “kill switches,” as part of the risk management framework. Despite its importance, the human-in-the-loop approach could introduce unintended risks. For instance, as payment flows are optimized as a whole, a payment delay caused by human approval could unintentionally increase liquidity risks and make risk-hedging strategies less effective (Bank of England 2026).

Although interruption mechanisms are well established in adjacent domains, such as algorithmic trading, these safeguards typically rely on regulated intermediaries, centralized control, and clear institutional accountability (International Organization of Securities Commissions 2015). Agentic payment scenarios may weaken these assumptions, particularly where AI systems operate across platforms or act under delegated authority, reducing the applicability of existing approaches. At the same time, interruption mechanisms are not without risk. If poorly designed, they may introduce single points of failure, expand the system’s attack surface, or enable denial of service through unauthorized or cascading activation (Russu 2025). Accordingly, kill switches should act as a layered governance capability embedded across the payment stack with distributed, rather than centralized,

<sup>9</sup> Human-in-the-loop mechanisms can mitigate certain risks associated with automated systems, but they are not sufficient on their own, particularly under time-compressed or clustered failure scenarios. This reinforces the importance of architectural and automated safeguards that limit reliance on real-time human intervention.

control. They should also include graduated responses rather than abrupt shutdowns, clear authority and auditability, and containment strategies that localize impact and avoid unnecessary system-wide disruption (Qin and others 2024).

Given the non-deterministic nature of generative models, agentic AI systems should not directly execute irreversible transactions without structural safeguards. The architectural separation of decision-making and execution layers, as proposed in previous sections, can act as an important mitigating factor. Agents in the decision layer propose or initiate an action, while a deterministic execution layer performs compliance checks or independent controls and execute the final transaction (Acharya 2025).

### **Private Sector Measures**

Payment networks are well advised on working toward becoming the trusted infrastructure layer. This includes launching agent-ready card products with embedded spending controls and real-time budget management; building global agent registries with identity verification, reputation scoring, and fraud monitoring; introducing dispute resolution frameworks tailored to AI-initiated transactions with clear liability allocation; and developing open protocol standards that work across various AI platforms to enable interoperability (Singh 2025). Digital wallet providers should deploy “agent-ready” capabilities, build agent friendly authentication using passkeys and biometric verification that works in conversational interfaces, and preserve merchant relationships ensuring that merchants remain “merchant of record” even when transactions originate through agents (Singh 2025).

For private sector participants, it is important to enable real-time access to clean, consistent data and map functions holistically to facilitate the integration of multiple types of AI, technological safeguards, and human review (RegTechLab 2025). Financial institutions and payment companies need to invest in training and recruitment to build teams capable of developing, implementing, and maintaining safe, robust, and interoperable AI systems (Uppuduri 2025). Finally, a robust cybersecurity posture is paramount when introducing agentic AI into payments. Autonomous AI agents often interact with multiple external systems, tools, and APIs, expanding the attack surface and exposing organizations to risks such as data exfiltration, tool misuse, and cross system-privilege escalation. Therefore, mitigation requires strong authentication, scoped authorization, and secure API governance when integrating AI agents with operational infrastructure (Errico, Ngiam, and Sojan 2025).

### **Public Sector Measures**

From a public sector perspective, agentic AI requires that regulators consider the implications of shifting from Know-Your-Customer to Know-Your-Agent requirements, where mandated verifiable identities for financial bots are linked to legal entities (WEF 2026). Traditional fraud models rely on human behavioral patterns, which become ineffective when transactions are initiated by autonomous agents. Hence, developing authentication frameworks that verify both the AI agent’s identity and the user’s delegated authority remains key. Some of the technologies described in Table 2 are designed to allow guardrails to enable auditability, logging, and explainability for payment agents to address some of the regulatory requirements.

As autonomous AI agents become integrated into financial and public systems, policymakers increasingly view them as part of critical digital infrastructure that requires robust governance and oversight. Therefore, ensuring safe deployment requires mechanisms that allow systems to operate autonomously while still enabling supervisory intervention when unsafe behavior emerges (Schmitz, Rystrøm, and Batzner 2025). Policymakers should deploy real-time monitoring systems to detect anomalies in agent behavior and transaction flows, maintain AI activity logs and audit trails, and generate automated alerts for abnormal agent behaviors (FSB 2025). Enabling real-time monitoring of risks requires appropriate data collection, incident reporting and disclosure policies, and AI-related supervisory skills to keep pace with technological progress. Regulators should allow testing of agentic payment systems in supervised environments, such as regulatory sandboxes, before full market deployment.

Countries have initiated AI governance frameworks, but more efforts are needed for regulatory coordination and harmonization to preempt potential spillover and contagion effects, as well as regulatory fragmentation, especially in cross-border payments. One notable example is Singapore which has introduced a model governance framework for agentic AI. The framework provides organizations with guidance on technical and nontechnical measures they need to put in place to deploy agents responsibly including a thorough ex-ante risk assessment by selecting appropriate agentic use cases and placing limits on agents' powers such as their autonomy and access to tools and data, reinforcing human accountability for agents by defining significant checkpoints at which human approval is required, implementing technical controls and processes throughout the agent lifecycle, and enabling end-user responsibility through transparency and education (IMDA 2026). AI-related regulation affecting agentic AI applications, such as EU's AI Act, is a key risk mitigation strategy that the public sector should consider (European Union Artificial Intelligence Act 2024).

## Conclusion

---

Agentic AI represents a potentially transformative development for payment systems. Software agents that interpret objectives, coordinate transactions, and interact autonomously across digital services can significantly expand the automation of financial activity. Early experimentation shows that agentic systems can improve the efficiency of e-commerce transactions, optimize cross-border payment routing, enhance liquidity management, and support real-time compliance monitoring.

However, these benefits arise alongside fundamental design tensions. Payment systems are built on deterministic infrastructures that guarantee predictability, accountability, and legal settlement finality.<sup>10</sup> Agentic AI systems, by contrast, rely on probabilistic reasoning and adaptive decision making. Therefore, integrating these technologies requires careful architectural design to ensure that automation does not undermine the core reliability of payment infrastructures.

The three-layer framework proposed in this note reconciles these differences through the separation of probabilistic decision making from deterministic authorization and settlement. This structure allows payment systems to leverage agentic technologies for improved orchestration and optimization while still maintaining robust safeguards around execution. In this model, innovation is concentrated upstream, where agents interpret intent and coordinate actions, while downstream layers preserve rules-based authorization and legally final settlement.

At the same time, several risks remain salient. Autonomous agents could misinterpret user intent, amplify market volatility through highly correlated actions, or create new vulnerabilities in cybersecurity, identity management, and liability frameworks. The emergence of machine-initiated transactions also challenges existing regulatory constructs built around human users, raising new questions around authentication, accountability, and dispute resolution.

Mitigating these risks will require coordinated action from both private and public stakeholders. Financial institutions must invest in governance structures, cybersecurity safeguards, and technical architectures that separate agentic reasoning from payment execution. Payment networks and technology providers will need to develop trusted identity frameworks and interoperable standards for Know-Your-Agent verification and delegated authority. At the same time, regulators may need to adapt supervisory approaches, including monitoring frameworks, testing environments, and governance standards for AI-mediated financial activity.

---

<sup>10</sup> The preservation of legal finality at the settlement layer, including for emerging technologies such as DLT, depends on explicit legal recognition and appropriate statutory safeguards.

In conclusion, the trajectory of agentic AI in payments will depend not only on technological capability but also on institutional design and governance choices. If implemented with appropriate safeguards, agentic AI could enhance efficiency, reduce friction in cross-border payments, and expand financial access. Without such safeguards, however, the same technologies could introduce new operational and systemic risks. Therefore, the challenge for policymakers and industry participants is not whether to adopt agentic technologies, but how to integrate them into payment systems in a way that preserves trust, stability, and accountability in an increasingly automated financial ecosystem.

## References

- Acharya, Vivek. 2025. "Secure Autonomous Agent Payments: Verifying Authenticity and Intent in a Trustless Environment." arXiv preprint arXiv:2511.15712. <https://arxiv.org/pdf/2511.15712>
- Aldasoro, Iñaki, Leonardo Gambacorta, Anton Korinek, Vatsala Shreeti, and Merlin Stein. 2024. "Intelligent Financial System: How AI Is Transforming Finance." BIS Working Papers No. 1194, Bank for International Settlements. <https://www.bis.org/publ/work1194.htm>
- Axelsen, Henrik, Valdemar Licht, and Jan Damsgaard. 2025. "Agentic AI for Financial Crime Compliance." arXiv preprint arXiv:2509.13137. <https://arxiv.org/pdf/2509.13137>
- Bank for International Settlements (BIS). 2024. "Annual Economic Report." <https://www.bis.org/publ/arpdf/ar2024e.htm>
- Bank for International Settlements (BIS). 2025. "The Use of Artificial Intelligence for Policy Purposes." Report submitted to the G20 Finance Ministers and Central Bank Governors. <https://www.bis.org/publ/othp100.pdf>
- Bank of England (BoE). 2022. "Artificial Intelligence and Machine Learning." Discussion Paper 5/22. <https://www.bankofengland.co.uk/prudential-regulation/publication/2022/october/artificial-intelligence>
- Bank of England (BoE). 2026. "Summary of AI Roundtables - February 2026." <https://www.bankofengland.co.uk/minutes/2026/february/summary-of-ai-roundtables-feb-2026>
- Beard, Nicolette V. 2025. "Ecommerce AI Agents: The Autonomous Technology Transforming Digital Retail." BigCommerce Blog. <https://www.bigcommerce.com/blog/ecommerce-ai-agents>
- Boston Consulting Group (BCG). 2025. "Agentic AI, Digital Currencies and Real-Time Transactions Reshape Global Payments Landscape." <https://www.bcg.com/press/22september2025-reshape-global-payments-landscape>
- Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond. 2023. "Generative AI at Work." NBER Working Paper No. 31161. [https://www.nber.org/system/files/working\\_papers/w31161/w31161.pdf](https://www.nber.org/system/files/working_papers/w31161/w31161.pdf)
- Cao, Mingxuan. 2026. "Bridging Information Asymmetry through AI-Driven Fintech: The Role of Digital Footprint Analytics in Financial Inclusion. International Business & Economics Studies Vol. 8, No. 1, 2026. <https://www.scholink.org/ojs/index.php/ibes/article/view/57059>
- Capco. 2025. "Agentic AI: Transforming Payments and Cash Management." <https://www.capco.com/intelligence/capco-intelligence/agentic-ai-transforming-payments-and-cash-management>
- Carichon, Florian, Aditi Khandelwal, Marylou Fauchard, and Golnoosh Farnadi. 2025. "The Coming Crisis of Multi-Agent Misalignment: AI Alignment Must Be a Dynamic and Social Process." arXiv preprint arXiv:2506.01080. <https://arxiv.org/pdf/2506.01080>
- Chiris, Lorenzo Satta, and Ayush Mishra. 2025. "AURA: An Agent Autonomy Risk Assessment Framework." arXiv preprint arXiv:2510.15739. <https://arxiv.org/abs/2510.15739>
- Convera. 2025. "How Agentic AI is Changing the Payments Landscape." <https://convera.com/blog/cross-border-payments/agentic-ai-in-payments>
- Desai, Akshar Prabhu. 2025. "Beyond the Subscription: Why Agentic Commerce Needs Stablecoins to Scale." Fintech Weekly.

<https://www.fintechweekly.com/magazine/articles/agentic-commerce-stablecoins-micropayments-ai-payments>

- Errico, Herman, Jiquan Ngiam, and Shanita Sojan. 2025. "Securing the Model Context Protocol (MCP): Risks, Controls, and Governance." arXiv preprint arXiv:2511.20920. <https://arxiv.org/abs/2511.20920>
- European Union Artificial Intelligence Act. 2024. <https://artificialintelligenceact.eu/>
- Financial Stability Board (FSB). 2022. "Assessment of Risks to Financial Stability from Crypto-Assets." <https://www.fsb.org/2022/02/assessment-of-risks-to-financial-stability-from-crypto-assets/>
- Financial Stability Board (FSB). 2023. "IMF-FSB Synthesis Paper: Policies for Crypto-Assets." <https://www.fsb.org/uploads/R070923-1.pdf>
- Financial Stability Board (FSB). 2025. "Monitoring Adoption of Artificial Intelligence and Related Vulnerabilities in the Financial Sector." <https://www.fsb.org/2025/10/monitoring-adoption-of-artificial-intelligence-and-related-vulnerabilities-in-the-financial-sector/>
- FinRegLab. 2025. "The Next Wave Arrives: Agentic AI in Financial Services." [https://finreglab.org/wp-content/uploads/2025/09/FinRegLab\\_09-04-2025\\_The-Next-Wave-Arrives-Main.pdf](https://finreglab.org/wp-content/uploads/2025/09/FinRegLab_09-04-2025_The-Next-Wave-Arrives-Main.pdf)
- Gartner. 2025. "Gartner Predicts Agentic AI Will Autonomously Resolve 80 Percent of Customer Service Issues without Human Intervention by 2029." <https://www.gartner.com/en/newsroom/press-releases/2025-03-05-gartner-predicts-agentic-ai-will-autonomously-resolve-80-percent-of-common-customer-service-issues-without-human-intervention-by-2029>
- Google. 2026. "Universal Commerce Protocol." Google Developers. <https://developers.google.com/merchant/ucp>
- Infocomm Media Development Authority. 2026. "Singapore Launches New Model AI Governance Framework for Agentic AI." <https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2026/new-model-ai-governance-framework-for-agentic-ai>
- Hogan Lovells. 2025. "Agentic AI in Financial Services: Regulatory and Legal Considerations." <http://hoganlovells.com/en/publications/agentic-ai-in-financial-services-regulatory-and-legal-considerations>
- Hu, Botao 'Amber', and Helena Rong. 2025. "Inter-Agent Trust Models: A Comparative Study of Brief, Claim, Proof, Stake, Reputation and Constraint in Agentic Web Protocol Design." arXiv preprint arXiv:2511.03434. <https://arxiv.org/abs/2511.03434>
- International Monetary Fund (IMF). 2025. "Risk and Resilience in the Global Foreign Exchange Market." Global Financial Stability Report. <https://www.imf.org/-/media/files/publications/gfsr/2025/october/english/ch2.pdf>
- International Organization of Securities Commissions (IOSCO). 2015. "Mechanisms for Trading Venues to Effectively Manage Electronic Trading Risks and Plans for Business Continuity." <https://www.iosco.org/library/pubdocs/pdf/ioscopd483.pdf>
- Joshi, Himanshu. 2025. "AI Governance by Design for Agentic Systems: A Framework for Responsible Development and Deployment." <https://storage.ghost.io/c/44/95/449506ca-034e-480f-9725-fcde08ef1cc1/content/files/2025/06/AI-Governance-by-Design-for-Agentic-Systems--A-Framework-for-Responsible-Development-and-Deployment.pdf>
- Kabadjova, Biliانا Alexandrova, Anton Badev, Saulo Benchimol Bastos, Evangelos Benos, Freddy CepedaLópez, James Chapman, Martin Diehl, Ioana Duca-Radu, Rodney Garratt, Ronald Heijmans, Anneke Kosse, Antoine Martin, Thomas Nellen, Thomas Nilsson, Jan Paulick, Andrei Pustelnikov,

Francisco Rivadeneyra, Mario Rubem do Coutto Bastos, and Sara Testi. 2023. “Intraday Liquidity around the World.” Bank for International Settlements Working Papers No 1089.  
<https://www.bis.org/publ/work1089.pdf>

Korinek, Anton, and Jai Vipra. 2025. “Concentrating Intelligence: Scaling Laws and Market Structure in Generative AI.” *Economic Policy* 40 (121), p. 4.

Mastercard. 2024. “Decision Intelligence – Beyond Mastercard.”  
<https://www.mastercard.com/content/dam/mccom/shared/business/b2b/reports/decision-intelligence-beyond-mastercard-playbook.pdf>

McKinsey & Company. 2025a. “How Agentic AI Can Change the Way Banks Fight Financial Crime.”  
<https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/how-agentic-ai-can-change-the-way-banks-fight-financial-crime>

McKinsey & Company. 2025b. “The Agentic Commerce Opportunity: How AI Agents are Ushering in a New Era for Consumers and Merchants.”  
<https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-agentic-commerce-opportunity-how-ai-agents-are-ushering-in-a-new-era-for-consumers-and-merchants>

McKinsey & Company. 2025c. “The End of Inertia: Agentic AI’s Disruption of Retail and SME Banking.”  
<https://www.mckinsey.com/industries/financial-services/our-insights/the-end-of-inertia-agentic-ais-disruption-of-retail-and-sme-banking>

Microsoft. 2025. “The 2025 Annual Work Trend Index: The Frontier Firm is Born.”  
<https://www.microsoft.com/en-us/worklab/work-trend-index/2025-the-year-the-frontier-firm-is-born>

National Institute of Standards and Technology (NIST). 2023. “Artificial Intelligence Risk Management Framework (AI RMF 1.0).” <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>

Ogbuonyalu, Uchenna Obiageli, Kehinde Abiodun, Selorm Dzamefe, Ezech Nwakaego Vera, Adewale Oyinlola, and Igba Emmanuel. 2024. “Assessing Artificial Intelligence Driven Algorithmic Trading Implications on Market Liquidity Risk and Financial Systemic Vulnerabilities.” *International Journal of Scientific Research and Modern Technology* 3 (4): 18–21.

Omar, Mahmud, Vera Sorin, Jeremy D. Collins, David Reich, Robert Freeman, Nicholas Gavin, Alexander Charney, Lisa Stump, Lisa Stump, Nicola Luigi Bragazzi, Girish Nadkarni, and Eyal Klang. 2025. “Multi-Model Assurance Analysis Showing Large Language Models are Highly Vulnerable to Adversarial Hallucination Attacks during Clinical Decision Support.” *Communications Medicine* 5: 330.  
<https://doi.org/10.1038/s43856-025-01021-3>

Onyekaonwu, Chinenye Blessing, Emmanuel Igba, and Amina Catherine Peter-Anyebe. 2024. “Agentic AI for Regulatory Intelligence: Designing Scalable Compliance Lifecycle Systems in Multinational Tech Enterprises.” *International Journal of Scientific Research and Modern Technology* 3 (12): 205–22.

Paypal. 2025. “PayPal Launches Agentic Commerce Services to Power AI-Driven Shopping.”  
<https://newsroom.paypal-corp.com/2025-10-28-PayPal-Launches-Agentic-Commerce-Services-to-Power-AI-Driven-Shopping>

Preis, Adam. 2026. “Agentic Finance: Building Trust in the Age of Autonomous Intelligence.” Finextra.  
<https://www.finextra.com/blogposting/30867/agentic-finance-building-trust-in-the-age-of-autonomous-intelligence?>

- Qin, Xingsheng, Frank Jiang, Chengzu Dong, and Robin Doss. 2024. "A Hybrid Cyber Defense Framework for Reconnaissance Attack in Industrial Control Systems." *Computers & Security* 136: 103506.
- RegTechLab. 2025. "The Next Wave Arrives: Agentic AI in Financial Services." Market Scan. [https://finreglab.org/wp-content/uploads/2025/09/FinRegLab\\_09-04-2025\\_The-Next-Wave-Arrives-Main.pdf](https://finreglab.org/wp-content/uploads/2025/09/FinRegLab_09-04-2025_The-Next-Wave-Arrives-Main.pdf)
- Reuters. 2025. "Citi, Ant International Pilot AI-Powered FX Tool for Clients to Help Cut Hedging Costs." <https://www.reuters.com/business/finance/citi-ant-international-pilot-ai-powered-fx-tool-clients-help-cut-hedging-costs-2025-07-18/>
- Russu, A. 2025. "Kill Switches in Connected Devices: The Emerging Risk Leaders Cannot Ignore." <https://www.resillion.com/blogs/kill-switches-in-connected-devices-the-emerging-risk-leaders-cannot-ignore/>
- Scalefocus. 2025. "The Role of AI in Cross-Border Payments." <https://www.scalefocus.com/blog/the-role-of-ai-in-cross-border-payments>
- Schmitz, Chris, Jonathan Rystrom, and Jan Batzner. 2025. "Oversight Structures for Agentic AI in Public-Sector Organizations." *Association for Computational Linguistics*. <https://arxiv.org/abs/2506.04836>
- Shivam. 2026. "What Does the India AI Impact Summit 2026 Signal for the Future of Payment Infrastructure." <https://www.pinelabs.com/blog/india-ai-impact-summit-payments>
- Shukanayev, Dastan. 2025. "Who Pays When the Agent Fails? Liability Frameworks for Autonomous Payment Systems in a Fragmented Regulatory Landscape." December 1. <https://ssrn.com/abstract=5864482>
- Singh, Anurag. 2025. "The Payments Revolution: How Agentic Commerce Will Transform B2C and B2B Transactions." <https://www.linkedin.com/pulse/payments-revolution-how-agentic-commerce-transform-b2c-anurag-singh-5l5tc/>
- Stripe. 2026. "Supporting Additional Payment Methods for Agentic Commerce." <https://stripe.com/blog/supporting-additional-payment-methods-for-agentic-commerce>
- Uppuduri, Vinod. 2025. "The Future of International Transactions: How AI is Transforming Cross-Border Payments." *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* 11 (1). <https://ijsrcseit.com/index.php/home/article/view/CSEIT251112157/CSEIT251112157>
- Visa. 2026. "Visa Protect A2A Payments." <https://corporate.visa.com/en/products/visa-protect-a2a-payments.html>
- World Economic Forum (WEF). 2025. "AI and Stablecoins: A Pairing for a More Intelligent Era of Online Business." <https://www.weforum.org/stories/2025/01/stablecoin-ai-business/>
- World Economic Forum (WEF). 2026. "AI agents could be worth \$236 billion by 2034 – if we get trust right." <https://www.weforum.org/stories/2026/01/ai-agents-trust/>
- Yan, Yineng, Xidong Wang, Jin Seng Cheng, Ran Hu, Wentao Guan, Nahid Farahmand, Hengte Lin, and Yue Li. 2025. "FaMA: LLM-Empowered Agentic Assistant for Consumer-to-Consumer Marketplace." <https://doi.org/10.48550/arXiv.2509.03890>



# PUBLICATIONS

**How Agentic AI Will Reshape Payments**  
**NOTE/2026/004**