European Journal of Risk Regulation (2026), **17**, 884–907
doi:10.1017/err.2026.10103

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

AI Agents in Payments: Applications, Risks and Regulations

David Restrepo Amariles[1], Damien Charlotin[1] and Liyun He-Guelton[2]

[1]Tax and Law, HEC Paris, France and [2]Independent scholar, France
Corresponding author: David Restrepo Amariles; Email: restrepo-amariles@hec.fr

Abstract

The integration of artificial intelligence (AI) agents into payment systems signals a profound shift in the architecture of financial transactions. Building on advances in large language models and autonomous systems, "agentic payments" refer to transactions initiated and completed by AI agents without direct human intervention. This article provides a conceptual and technical analysis of agent-enabled payment systems, examining their operational logic, defining features and emerging use cases across retail, e-commerce and decentralised finance. It distinguishes agentic payments from traditional automated systems by emphasising autonomy, contextual reasoning and adaptability. The article further identifies and categorises a range of technical, legal and societal risks, including cybersecurity vulnerabilities, liability gaps, regulatory non-compliance, and potential economic disruption. Through case studies and architectural illustrations, it highlights both the innovation potential and governance challenges posed by agentic systems. It argues that current regulatory frameworks – designed for human-intermediated payments – are ill-equipped to address the dynamic and decentralised nature of agent-led transactions. The article concludes by proposing a multi-layered governance framework combining core regulatory requirements with supporting ecosystem measures to ensure accountability, security, and transparency in the age of autonomous financial agency.

Keywords: agents; AI regulation; payments

I. Introduction

The renewed interest in artificial intelligence (AI) agents, driven in part by the rapid advancement and widespread adoption of large language models (LLMs), marks a significant development in the trajectory of AI technologies. AI agents are increasingly viewed as a promising architecture, with the capacity to autonomously execute complex tasks in ways that may reshape both institutional workflows and individual user experiences. These agents are designed to make autonomous decisions within the bounds of pre-defined goals or instructions, and their practical utility will depend not only on the sophistication of their underlying models, but also on the breadth of tools and systems to which they are granted access. Among these, the integration of payment functionalities appears increasingly inevitable within emerging retail agentic workflows, with recent examples including Google's and Amazon's *buy for me*[1] applications.

[1] The "buy for me" functionality enables Google and Amazon to automatically search for and purchase goods online under the supervision of the customer. For further details on Google: <https://blog.google/products/shopping/google-shopping-ai-mode-virtual-try-on-update/> (last accessed 7 April 2026). Further details on

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/err.2026.10103

This makes sense: enabling AI agents to initiate and complete transactions could reduce operational friction and expand payment optionality, to the benefit of the customer. However, such integration also introduces significant challenges, particularly in relation to accountability, regulatory compliance, financial integrity, and security.

This paper examines the emergence of payment-enabled AI agents and explores the technological developments underpinning this evolution, focusing on both technology and payment providers (A). We then propose a definition of "agentic payments"; and describe in particular how such payments will differ from the output of mere automated systems (B), before engaging in a review of the risks involved in such systems, including legal, technical, ethical, and economic challenges (C). Building on this, we critically assess whether the current regulatory framework is sufficiently robust to address these risks and govern these evolving technologies effectively (D). Finally, we conclude by outlining key guardrails necessary to ensure the responsible deployment and governance of agentic payment systems (E).

II. The development of payment-enabled agents

While the ambition to develop autonomous AI agents has long been a feature of broader artificial intelligence research,[2] progress in this area has been relatively incremental, and adoption in real-world contexts remains limited. A 2022 review of existing chatbot systems, for example, underscored the "weak appropriateness" of most responses generated by agent-based applications, particularly those reliant on rule-based approaches.[3] These limitations have contributed to the comparatively slower integration of agentic AI in commercial and professional settings, especially when contrasted with the more widespread adoption of machine learning techniques across sectors.[4]

Yet, the advent of LLMs, particularly their general-purpose capacity to engage in contextual reasoning, has reinvigorated interest in agents. This renewed attention stems in part from the recognition that agents may serve as a viable mechanism to mitigate some of the intrinsic limitations of LLMs – most notably their lack of persistent memory, limited capacity for long-term planning, and difficulty in executing complex, multi-step tasks.[5] Crucially, while LLMs function as monolithic tools, agents represent modular systems capable of invoking and coordinating multiple tools, including both AI-driven and conventional software components. Consequently, there is growing momentum within industry to develop agentic architectures capable of automating entire workflows. These systems, often referred to as Agentic AI, increasingly rely on multi-agent orchestration, wherein collections of agents collaborate to perform structured, interdependent operations with a level of autonomy and scalability unattainable by standalone models.[6]

Amazon's functionality, which enables purchases outside of merchants listed on its own marketplace, see: <https://www.aboutamazon.com/news/retail/amazon-shopping-app-buy-for-me-brands> (last accessed 7 April 2026).

[2] S Albrecht and P Stone, "Autonomous Agents Modelling Other Agents: A Comprehensive Survey and Open Problems" (2018) 258 Artificial Intelligence 66. The term "agent" was introduced into computer science in the 1970s to designate autonomous software entities acting on behalf of a user or another program. In this article, the word "agent" will be used for the more recent concept of an "AI agent."

[3] B Luo, R YK Lau, C Li and Y-W Si, "A Critical Review of State-of-the-Art Chatbot Designs and Applications" (2022) 12(1) Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery e1434.

[4] TH Davenport and R Ronanki, "Artificial Intelligence for the Real World" (2018) 108 Harvard Business Review 112.

[5] L Wang and others, "A Survey on Large Language Model Based Autonomous Agents" (2024) 18 Frontiers of Computer Science 186345.

[6] R Sapkota, KI Roumeliotis and M Karkee. "Ai Agents vs. Agentic AI: A Conceptual Taxonomy, Applications and Challenge" (2025) *arXiv preprint* arXiv:2505.10468.

https://doi.org/10.1017/err.2026.10103

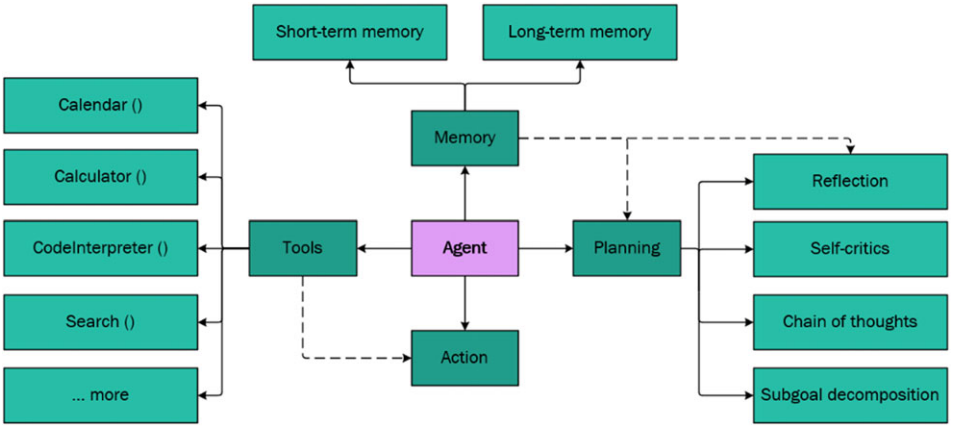


Figure 1. High-level architecture of an advanced AI agent.

In what follows, we define "AI agents" as any system that, when prompted with a request, (i) devises a strategy to answer the request in view of the goals and constraints; and (ii) successfully implements that strategy.[7] This differs from a simple call to an LLM equipped with a tool, in that the agent is meant to iterate[8] between (i) and (ii) until the request has been fully and satisfactorily answered.

1. The state of the art in technology

At the forefront of the developments described in this article is the modularisation of agent functionalities, a strategy that operationalises agents by equipping them with purpose-specific tools and modules. This approach allows agents to combine their general, LLM-based reasoning abilities, with actionable tools for completing specific tasks. Figure 1 presents a high-level architecture of an advanced AI agent, integrating planning, memory, tool use and autonomous action. At the core is the agent, which interacts with a memory system (short- and long-term), a planning module, and various external tools (e.g., calculator, search, calendar). The planning module coordinates higher-order reasoning functions, including reflection, self-critique, chain of thought, and subgoal decomposition, enabling the agent to structure tasks, revise plans, and adapt strategies. The agent executes actions based on tool outputs and planning inputs, while memory supports persistence and learning. This architecture enables complex, goal-directed behavior and highlights key features of agentic AI: autonomy, reasoning, and tool integration.

From this architectural basis, agents have demonstrated impressive capabilities in solving a wide array of tasks.[9] A modular agent, for example, may feature an "information retrieval" module to gather additional data which is aggregated with raw user input, a

[7] Other definitions exist for different purposes. See, generally, C Huyen, *AI Engineering: Building Applications with Foundation Models* (O'Reilly Media 2025) <https://huyenchip.com/2025/01/07/agents.html> (last accessed 7 April 2026).

[8] The emphasis on iteration is common to the definition of agents, including in transactional contexts: see D Stern and D Greenwood, "From Fine Print to Machine Code: How AI Agents are Rewriting the Rules of Engagement" (*Codex Blog*, 14 January 2025), <https://law.stanford.edu/2025/01/14/from-fine-print-to-machine-code-how-ai-agents-are-rewriting-the-rules-of-engagement/> (last accessed 7 April 2026).

[9] S Schmidgall and others, "Agent Laboratory: Using LLM Agents as Research Assistants" (*arXiv*, 8 January 2025) <https://arxiv.org/abs/2501.04227> (last accessed 9 January 2025), p. 3, and the literature cited.

https://doi.org/10.1017/err.2026.10103

"Reasoning" module to analyze inputs and select a course of action, "tool assessment" to identify the most relevant tools for a given action and use them optimally and an "action " module designed to translate the agent's decisions into specific outputs.[10] The back and forth between modules, their input and outputs, is what allows the agent to iteratively refine its approach to completion of the task. A similar approach consists in using several agents with well-defined tasks.[11] Regardless of the architecture, the point is to promote flexibility in the achievement of a task, by matching reasoning and action in a self-reinforcing manner that relies on a succession of reasoning/action steps.[12]

In parallel, agents are increasingly able to interact with external systems, particularly through Application Programming Interfaces (APIs).[13] APIs serve as gateways, allowing agents to interface seamlessly with financial platforms, services and data sources. For instance, agents can query financial APIs to retrieve real-time data – such as exchange rates, transaction histories or account balances – and use this information to autonomously make payment-related decisions.

The growing sophistication of these integrations has already enabled basic autonomous transaction capabilities, particularly in controlled environments such as algorithmic trading systems or smart-contract-driven payments in decentralised finance. These systems demonstrate that agents can execute pre-programmed financial instructions, albeit still within rigid frameworks.

Within this architecture, the introduction of a *payment module or agent* becomes not only plausible but it might also soon be a requirement for the attractiveness of many applications. Such a module would enable agents to initiate and process payments autonomously, effectively bridging the gap between reasoning and action in financial contexts. Early indicators of this evolution can already be observed in technologies like programmable wallets, smart contracts and embedded payment systems that allow autonomous triggers based on agent-driven inputs.

It is important to situate agentic systems within broader enterprise trends, notably *hyperautomation*, which aims to combine robotic process automation, process mining, AI and business rules into integrated workflows. Agentic payments do not compete with such automation architectures, but they complement them. Agents are particularly well-suited for handling open-ended tasks and edge-case reasoning within larger automated pipelines, serving as modular, intelligent nodes capable of context-aware action where conventional automation falls short. As such, they may be deployed selectively within regulated domains that require both flexibility and auditability.

2. *The payment ecosystem as fruitful ground for agentic payments*

These technological developments have paved the way for the expectation that agents will soon become customers,[14] or otherwise deploy or receive financial value. Recognising this trajectory, several key actors in the payment ecosystem have begun adapting to accommodate agentic functionality. These efforts reflect a dual recognition of opportunity:

[10] Ibid., p. 2.

[11] Y Talebirad and A Nadiri, "Multi-Agent Collaboration: Harnessing the Power of Intelligent LLM Agents" (2023) arXiv preprint arXiv:2306.03314 <https://arxiv.org/abs/2306.03314> (last accessed 7 April 2026).

[12] See Anthropic, "Building Effective Agents" (19 December 2024), <https://www.anthropic.com/research/building-effective-agents> (last accessed 7 April 2026), for different examples of agentic patterns, including the Parallelisation, Orchestrator, Evaluator workflows.

[13] SG Patil, T Zhang, X Wang and JE Gonzalez, "Gorilla: Large Language Model Connected with Massive APIs" (2023) arXiv preprint arXiv:2305.15334 <https://arxiv.org/abs/2305.15334> (last accessed 7 April 2026).

[14] Gartner, "Gartner Says Machine Customers Represent One of the Biggest New Growth Opportunities of the Decade" (16 March 2023).

https://doi.org/10.1017/err.2026.10103

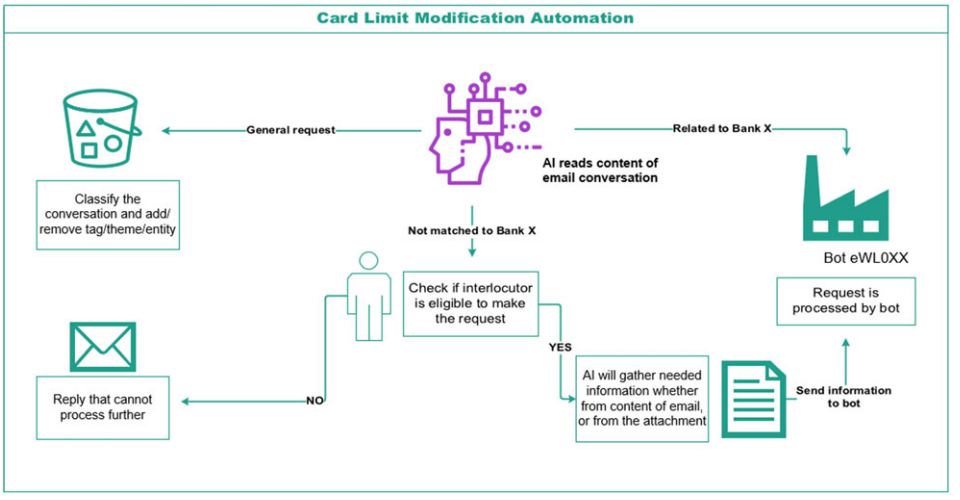


Figure 2. Agentic system for credit card limit modification.

greater operational efficiency for providers and an enhanced, frictionless experience for end users.

For instance, some payment providers acknowledge that an "LLMs engineering strategy is crucial in the payment sector,"[15] contemplating disruptions on both the producer (greater efficiency, AI agents) and consumers (enhanced shopping experience) sides. The Boston Consulting Group echoed this sentiment in a dedicated analysis,[16] highlighting that payment system providers, characterised by "high-volume, high-frequency, data-rich operations that involve human-intensive workloads,"[17] are particularly well-suited for innovation through LLM-based agents. These studies reflect a growing consensus that payment infrastructures must evolve to meet the needs of actors capable of autonomously initiating, approving or managing transactions.

A concrete example is the implementation of AI agents to automatically manage the modification of credit card limits, as patented by Worldline. Figure 2 illustrates Worldline's AI agent automating the workflow for card limit modification operating within a semi-autonomous communication and transaction system. The agent functions as a mediating interface between customer inquiries and backend financial processing infrastructure, exemplifying the integration of natural language processing (NLP), eligibility verification, and robotic process automation (RPA) in a financial compliance context.

In this figure, upon receiving an email request, the AI agent initiates its operation by parsing the content of the email conversation. It then determines whether the query pertains to a specific institution (referred to as "Bank X"). If the content is not matched to Bank X, the request is redirected into a general classification module. There, the agent assigns or removes thematic tags and entities and provides a standardised response indicating that the request cannot be processed further – effectively performing both automated triage and message classification.

If the content is related to Bank X, the agent proceeds to verify whether the interlocutor is authorised to make such a request. This eligibility assessment introduces a

[15] Worldline, "Strategizing large Language Model integration in the payment ecosystem" (3 April 2024) <https://worldline.com/en/home/main-navigation/resources/resources-hub/blogs/2024/strategizing-large-language-model-integration-in-the-payment-ecosystem> (last accessed 7 April 2026).

[16] BCG, "Finding the Sweet Spot for Generative AI in Payments" (7 December 2023).

[17] To these features could be added (i) stringent security requirements; and (ii) need for real-time processing.

https://doi.org/10.1017/err.2026.10103

rule-based evaluation step aligned with compliance protocols. If the requester is found to be ineligible, a rejection message is generated. Conversely, if eligibility is confirmed, the agent autonomously extracts relevant information – either from the body of the email or from attached documents – required to fulfill the card limit modification request.

Once the required data has been extracted, the information is transferred to an RPA component (Bot eWL0XX), which is responsible for processing the request. This final hand-off represents a seamless handover between AI-enabled front-end systems and enterprise back-end automation infrastructure, closing the loop between human communication and automated execution.

This workflow demonstrates a modular and conditional agentic design, combining document analysis, compliance logic, task delegation, and structured decision-making. The model also exemplifies how AI agents can extend traditional workflow automation by engaging in context-sensitive dialogue analysis, offering a foundation for more advanced implementations in banking and payment domains.

Other concrete initiatives are also flourishing in the sector. In November 2024, for instance, Stripe introduced an "agent toolkit," designed to allow agents to help earn and spend funds, facilitate common support operations, and bill for usage with metered billing.[18] Stripe has since gone much further. In February 2026, Stripe launched a preview of "machine payments," integrating the x402 protocol (an open standard developed by Coinbase that repurposes the HTTP 402 "Payment Required" status code) to allow developers to bill AI agents directly using USDC stablecoins on the Base network.[19] Then in March 2026, Stripe and Tempo (a payment-focused blockchain) launched the Machine Payments Protocol (MPP),[20] an open, rail-agnostic standard for agent-to-service payments – with Visa as a design partner extending MPP to card-based payments.[21] This shows the ecosystem has moved from the initial toolkits to purpose-built financial infrastructure for agents.

Meanwhile, start-ups are emerging based on the expectation that 'tomorrow's work will not only be paid by humans,[22] anticipating a future in which AI agents may engage human freelancers to perform tasks beyond their operational scope, or be granted direct access to financial resources to execute such tasks autonomously.[23]

This evolution is also rapidly being embraced in e-commerce, with major digital platforms increasingly reconfiguring transactional architectures around AI agents. Google's recent introduction of "shop with AI mode,"[24] supported by its Gemini 2.5 model, marks a significant shift in the structure of e-commerce. This feature enables users to complete purchases directly through Google Search, thereby bypassing traditional e-commerce interfaces such as product pages, shopping carts and external retailer websites. When combined with virtual try-on functionalities and conversational interfaces enabled by generative AI, this development constitutes a substantial move toward an

[18] See S Kaliski, "Adding payments to your LLM agentic workflows" (14 November 2024), <https://stripe.dev/blog/adding-payments-to-your-agentic-workflows> (last accessed 7 April 2026).

[19] B Danga, "Stripe adds x402 integration for USDC agent payments on Base" (*The Block*, 11 February 2026), <https://www.theblock.co/post/389352/stripe-adds-x402-integration-usdc-agent-payments> (last accessed 7 April 2026). See also Stripe, "Machine payments" (2026), <https://docs.stripe.com/payments/machine> (last accessed 7 April 2026).

[20] J Weinstein and S Kaliski, "Introducing the Machine Payments Protocol" (*Stripe Blog*, 18 March 2026), <https://stripe.com/blog/machine-payments-protocol> (last accessed 7 April 2026).

[21] PYMNTS, "Visa Scales Agentic Commerce Through Stripe Protocol Collaboration" (18 March 2026), <https://www.pymnts.com/visa/2026/visa-scales-agentic-commerce-through-stripe-protocol-collaboration/> (last accessed 7 April 2026).

[22] PaymanAI, <https://www.paymanai.com/> (last accessed 7 April 2026).

[23] Skyfire, <https://skyfire.xyz/> (last accessed 7 April 2026).

[24] Google, Shop with AI <https://blog.google/products/shopping/google-shopping-ai-mode-virtual-try-on-update/> (last accessed 7 April 2026).

https://doi.org/10.1017/err.2026.10103

uninterrupted, agent-led purchasing experience that integrates product discovery and transaction within a single interaction layer.

This reconfiguration has implications that extend beyond interface design. By collapsing multiple stages of the transaction process, Google's approach effectively disintermediates several key segments of the existing e-commerce infrastructure, including payment processing, checkout optimisation and fraud prevention. Control over the consumer experience is thereby shifted away from merchants and brand websites toward platform-owned AI agents. Traditional digital marketing strategies such as search engine optimisation (SEO), landing page optimisation, and brand engagement are rendered less effective as users increasingly interact with AI intermediaries rather than directly with vendors. In this context, structured product data, optimised for machine-readable formats, becomes a critical asset.

Amazon, by contrast, is pursuing a more integrated yet comparatively conservative approach to agentic commerce. Its "Buy with Me"[25] feature, embedded within the Amazon shopping app, allows users to discover and purchase products from third-party vendors not directly listed on Amazon's marketplace. The company's AI assistant can complete transactions on external retailer websites while Amazon handles the transmission of payment credentials and customer data, thereby maintaining transactional visibility. Importantly, fulfillment and customer service responsibilities remain with the original retailer, preserving a degree of brand autonomy within the broader platform architecture.

While both Google and Amazon position AI agents at the centre of the consumer journey, their approaches diverge in the extent to which they reconfigure existing commercial relationships. Google's model suggests a more disruptive paradigm, displacing traditional web-based commerce structures and redirecting consumer engagement toward platform-owned intermediaries. Amazon's strategy, in contrast, extends platform functionality without fully dismantling the established roles of third-party retailers. Taken together, these developments illustrate a broader shift in the digital economy, as transactional authority migrates from websites to autonomous agents capable of initiating and managing purchases across decentralised ecosystems.

In parallel, the crypto-financial sector has moved at a particularly accelerated pace, driven by its inherent alignment with agentic payment principles: smart contracts, tokenisation, and digital assets are all capable of enabling agents to exchange value seamlessly within decentralised systems. This is why Coinbase, for example, introduced AI-driven crypto wallets, explicitly addressing the gap in traditional banking by stating that "AI agents cannot get bank accounts, but they can get crypto wallets."[26] Notably, Coinbase also developed the x402 protocol, an open payment standard that repurposes the previously dormant HTTP 402 ("Payment Required") status code to enable on-chain payments directly within standard web requests – a protocol since adopted by Stripe for its own machine payments infrastructure.[27]

These developments demonstrate that the payment ecosystem, encompassing both traditional financial institutions and decentralised finance technologies, is actively preparing for the emergence of agent-mediated transactions. This evolution involves not only a reassessment of conventional business models of payment providers, but also the development and implementation of governance frameworks designed to identify and mitigate the legal, operational and security risks inherent in this new transactional infrastructure.[28]

[25] Amazon, Buy for Me, <https://www.aboutamazon.com/news/retail/amazon-shopping-app-buy-for-me-brands> (last accessed 7 April 2026).

[26] "AI wallets," <https://docs.cdp.coinbase.com/> (last accessed 7 April 2026).

[27] "Welcome to x402," <https://docs.cdp.coinbase.com/x402/welcome> (last accessed 7 April 2026).

[28] A Gerety, "AI agents have brains, but where are their wallets?" (*QED Investors*, 10 October 2024).

https://doi.org/10.1017/err.2026.10103

III. Agentic payments

In what follows, we define agentic payments as payments made autonomously by AI-driven agents, acting within predefined parameters but without direct human intervention.

This definition highlights the shift from human-dependent payment systems to agents capable of initiating, managing, and completing transactions. To fully grasp the significance of this concept, we will elaborate on its core features (a), distinguish agentic payments from existing automated systems (b), and illustrate the definition with relevant examples (c).

1. Core features of agentic payments

Agentic payments are characterised by three key features: autonomy, contextual reasoning and adaptability.[29]

Autonomy refers to the agent's ability to initiate and execute payments without immediate human oversight. Unlike traditional automated systems that follow rigid, pre-programmed instructions, agentic payments involve agents acting independently within predefined boundaries, such as budgetary constraints or task-specific goals. For instance, an agent could be programmed to make recurring payments, microtransactions or contextual buying decisions based on real-time data.

Contextual reasoning reflects the agent's capacity to make payment-related decisions informed by broader inputs from its environment. This includes, for instance, adjusting to dynamic pricing in real-time, evaluating contractual obligations, or assessing usage-based billing conditions before initiating a payment. Optimally, agentic decision-making processes should extend beyond a simple "if-then" logic, incorporating reasoning derived from external, and often evolving, data and factors.

Adaptability allows agents to adjust their behavior based on real-time feedback or changing objectives. If, for example, a payment fails due to insufficient funds, an adaptive agent might renegotiate terms, delay execution or explore alternative payment pathways. The key consideration here is that agents maintain a certain amount of control over their approach to executing a task.

In turn, these features will define to varying degrees a new ecosystem of payment agents, which will include, at the very least:

- Human-delegated agents acting on behalf of consumers,
- Merchant-deployed agents optimising sales and customer experience, and
- Bank-side agents enhancing security and transaction efficiency.

This multi-agent environment will fundamentally transform how payments are initiated, processed, and secured.

2. How they differ from existing automated systems

AI agents are not the first instance of payments occurring without human intervention. Ever since societies transitioned from purely cash-settled transactions to computer-assisted ones, automated payments have become a regular fixture in both corporate and personal financial operations.

[29] These features match older definitions of agents, that have insisted, e.g., on "autonomy, social abilities, reactivity, and proactiveness": see, e.g., SMC Loureiro, J Guerreiro and I Tussyadiah, "Artificial Intelligence in Business: State of the Art and Future Research Agenda" (2021) 129 Journal of Business Research 911.

https://doi.org/10.1017/err.2026.10103

This includes, for example, standing orders and programmatic trading, which execute payments or trades strictly in accordance with pre-programmed rules, with no capacity to adjust dynamically to changing conditions; or payment protocols, such as SEPA orders or recurring subscriptions, which replicate a single instruction at defined intervals following an initial authorisation. In both these cases, the payment system does not evaluate real-time changes or respond to external inputs beyond the original directive.

The key distinction lies in agents' capacity to dynamically choose actions based on evolving parameters or unforeseen circumstances. Unlike automated systems, which operate deterministically, agents can adjust their decisions in real time, balancing multiple inputs and goals. For instance, while an automated subscription payment simply executes a pre-set withdrawal each month, an agent system managing cloud computing costs could monitor usage, market rates, and budgetary constraints to autonomously decide on a payment amount, trigger a renegotiation request, or delay execution. Another crucial distinction is the ability of agents to engage in natural human interaction. Agents can converse with users in a human-like manner, allowing people to place, modify or cancel orders at any time through intuitive dialogue. This interactive capability enhances flexibility and user control in ways traditional automated systems cannot. Put another way, agents represent a middle way between a full spelling out of the components of a given task (which may stumble on out-of-distribution situations) and the flexible, often intuitive approach of human beings to such tasks.[30]

This distinction does not imply that agentic systems and automated systems cannot, in some instances, achieve the same end result. Rather, they represent two distinct technological approaches to similar problems: automated systems excel in fixed, predictable scenarios, while agentic systems thrive in dynamic environments requiring real-time decision-making. And indeed, for a wide array of problems agents might add unnecessary complexity.[31]

In other words, the emergence of agentic payments signals a shift from deterministic execution to situationally adaptive processes, with the potential of transforming how transactions are managed and optimised. In particular, and beyond technical improvements, AI agents may substantially elevate the user experience across both the payment process and the broader purchasing journey, creating more intuitive, personalised interactions.

3. *Illustrative examples*

It is possible to identify several emerging and existing implementations that offer insights into the operational potential of AI agents in payments. Consider, for instance, a scenario involving a customer-assistance agent tasked with identifying a flight or train journey under specific parameters, such as price ceilings, time constraints, or layover preferences. The agent would continuously monitor booking resources, dynamically balance conflicting priorities (e.g., cost versus convenience), and autonomously process payment once an "optimal" option is found.

Agents can also be deployed to minimise transaction risks during payments. For instance, Visa said its deployment of AI in this field already helped to "prevent fraud amounting to an estimated $27 billion in 2022."[32] A particularly noteworthy application is

[30] Another distinction should be made between agents and chatbots: while they may manifest through chat interfaces, agents are not designed for conversation *per se*, but for delegated task execution within constraints.

[31] See Anthropic, "Building effective agents" (19 December 2024), <https://www.anthropic.com/research/building-effective-agents> (last accessed 7 April 2026).

[32] S Rodchua, SC Korir, and SW Ogwangi, "Artificial Intelligence and its Impact on Management of Network Security and Cybercrime" (2024) Viva Tech Vegas: Advancing Technology, Management and Applied Engineering, ATMAE Conference Proceedings.

https://doi.org/10.1017/err.2026.10103

the Autonomous Transaction Risk Analysis (ATRA) system patented by Worldline, which acts as a supportive agent layer between customer-facing agents and the underlying payment infrastructure. ATRA agents autonomously evaluate two axes of risk before any transaction is authorised: the trustworthiness of the agent initiating the transaction (based on prior behavior, transaction history and interactions) and the reliability of the merchant, incorporating reputation data and transactional patterns. Where risks are minimal, agents enjoy full autonomy to complete transactions instantly. Conversely, where the ATRA score is elevated – due to anomalies such as unusual purchase types, irregular transaction volumes, or novel contexts – the system either escalates to human intervention or outright blocks the payment.[33] This kind of dual-agent verification introduces a dynamic, real-time layer of risk control, effectively enabling a scalable trust architecture for agentic payments.

Beyond risk mitigation, agentic micropayments are already operational in certain contexts. For instance, following Stripe's integration of the x402 protocol, CoinGecko launched x402-powered API endpoints allowing autonomous agents to access real-time price and on-chain data for a flat rate of $0.01 USDC per request, without requiring an account or API key.[34] This represents one of the first live instances of fully autonomous, pay-per-use agent commerce at scale.

These use cases illustrate the potential of agentic systems to enhance security while maintaining payment efficiency.

IV. Risks of AI agents

As mentioned above, actors in the payment ecosystem are gearing up for the emergence of agentic payments not only with a view to seizing an opportunity, but also because the payment infrastructure may be ill-prepared for that change of paradigm. Notably, a large part of the security infrastructure in this field focuses on making sure that humans, not automated systems, are triggering a request for payment.[35] In the same vein, many risk mitigation systems are now based on thresholds based on typical human behaviors. Cataloguing these risks can be done on a spectrum from technical and operational to legal and societal.

1. Technical and operational risks

One of the most immediate concerns is autonomy failures, where agents might make unintended or harmful payment decisions due to flawed programming or inadequate data inputs. Unlike automated systems that operate deterministically, agents can engage in reasoning that may lead to unpredictable outcomes. A common issue in automated systems, such as becoming stuck in loops or faltering due to random glitches, is magnified in agentic payments because of the agents' higher complexity and decision-making independence. The recursive character of agents, which ensures their flexibility, is also a potential liability: if a step in an agent's stack of actions has but a 1% chance of failing, the

[33] See Worldline, "Strategizing large Language Model integration in the payment ecosystem" (3 April 2024) <https://worldline.com/en/home/main-navigation/resources/resources-hub/blogs/2024/strategizing-large-language-model-integration-in-the-payment-ecosystem> (last accessed 7 April 2026).

[34] B Danga, "Stripe adds x402 integration for USDC agent payments on Base" (*The Block*, 11 February 2026), <https://www.theblock.co/post/389352/stripe-adds-x402-integration-usdc-agent-payments> (last accessed 7 April 2026); see also CoinGecko, "Instant Access to Crypto Prices & Market Data with x402", <https://www.coingecko.com/en/api/x402> (last accessed 7 April 2026).

[35] A Gerety, "AI agents have brains, but where are their wallets?" (*QED Investors*, 10 October 2024): "In fact, one of the most consistent themes in fraud detection today is trying to determine when an automated system is presenting itself with human credentials."

https://doi.org/10.1017/err.2026.10103

odds keep climbing as the agent comes back to this step.[36] For example, an agent managing cloud resources might incorrectly assess pricing trends and inadvertently overspend, causing financial loss.

Relatedly, recent experimentations have highlighted the extreme variance in agentic performance, even across repeated identical setups.[37] The best models oscillated between outperforming humans and catastrophic failure, e.g., complete business collapse due to misinterpreting delivery timing. For regulators and payment providers, this suggests that statistical average performance is not a sufficient safety metric. Robustness and predictability across runs matter far more in financial contexts.

Another critical risk is cybersecurity vulnerabilities, as agentic payment systems introduce new attack surfaces. The increased interaction of agents with payment platforms creates additional opportunities for hacking, spoofing and unauthorised transactions. Attackers may exploit an agent's decision-making processes, redirecting payments or manipulating the data that informs an agent's choices. A real-life example of this process was provided by Freysa AI, a cryptocurrency challenge in which users were incentivised to bypass the protection of an AI agent that guarded a sum of money with instructions not to wire it. While each prompt cost users a fixed amount, users could win the accumulated amounts by convincing the agent to transfer them – which happened after the agent gathered around 47,000 USD.[38]

Interoperability challenges also pose significant hurdles, as integrating agentic payment systems with legacy infrastructure (ACH, wire transfers, card networks, etc.) or even newer fintech platforms can prove difficult.[39] Payment systems today are often fragmented, with varying degrees of technological sophistication. Ensuring that agents can operate seamlessly across different systems while maintaining security and efficiency requires rethinking how payment infrastructures are designed. For example, existing platforms may lack the ability to recognise or validate agentic payment requests, leading to transaction failures or delays.

From a technical standpoint, managing these risks demands a fundamental shift in approach. The current paradigm, which largely distinguishes between human and bot interactions, must evolve toward frameworks that validate payments based on their legitimacy, regardless of whether they are initiated by a human or an agent. This shift requires redefining fraud detection, authentication and error mitigation protocols. For instance, instead of relying on behavioral norms tied to human users, future systems might assess the logic and context behind an agent's actions to determine whether a payment is valid. This reorientation will be critical in ensuring the secure and efficient operation of agentic payment systems.

2. *Legal and regulatory risks*

Agentic payments introduce significant legal and regulatory challenges, many of which stem from the autonomous nature of agents and their interactions with existing payment systems and frameworks. These risks highlight the urgent need for regulatory adaptation

[36] C Huyen, AI Engineering: Building Applications with Foundation Models (O'Reilly Media 2025), <https://huyenchip.com/2025/01/07/agents.html> (last accessed 7 April 2026).

[37] S Datta, K Lin, N Lambert and T Wolf, "Vending-Bench: A Benchmark for Long-Term Coherence of Autonomous Agents" (2025) arXiv preprint arXiv:2502.15840 <https://arxiv.org/abs/2502.15840> (last accessed 7 April 2026).

[38] M McColl, "If you can make this AI bot fall in love, you could win thousands of dollars" (*TechCrunch*, 6 December 2024).

[39] R Loganathan, "Future of Payments: AI Agents in Banking" (9 October 2024) Sardine AI <https://www.sardine.ai/blog/ai-agents-payments> (last accessed 7 April 2026).

https://doi.org/10.1017/err.2026.10103

and oversight to address gaps that could undermine trust and security in this emerging paradigm.

A primary concern is accountability and liability. When agents make errors, whether through flawed programming, misinterpretation of data, or external manipulation, questions arise about who bears responsibility for the consequences. For instance, if an agent authorises an unauthorised transaction or engages in an unlawful activity, should liability rest with the agent's creator, owner, or the payment platform facilitating the transaction? Current legal frameworks lack clarity on such matters, leaving stakeholders exposed to legal and financial risks. To some extent, these issues can be settled on the basis of the existing legal principles governing principal-agent relationships,[40] but this framework might prove inadequate to cover these new relationships. Likewise, existing rules about automatic payments[41] are not necessarily geared to the autonomous character of agents.

Compliance gaps are also critical. Payment systems are heavily regulated, particularly in areas such as anti-money laundering (AML) and know-your-customer (KYC) requirements. As the Boston Consulting Group notes, "[e]ven with GenAI, however, some things will not change. High regulatory and security protocols will remain, and merchants will still have know-your-customer obligations."[42] Agents, however, complicate compliance processes by operating autonomously and potentially in real time. Ensuring that agents meet these obligations without human oversight demands significant technological advancements and regulatory adjustments.

The access to and handling of personal data present another critical issue. For agents to function effectively in payment contexts, they will require access to extensive personal and financial information, tethering their decisions to an individual's preferences and circumstances. However, this access exposes agents to heightened risks of data leakage, particularly through vulnerabilities such as prompt injection attacks.

Systems like Google's Mariner, which autonomously prepares to purchase ingredients online for recipes, illustrate this challenge. Google itself acknowledged the difficulty of balancing agent autonomy with human control, particularly for sensitive actions like payments.[43] This issue ties into what has been dubbed the "challenge of gullibility," where agents are susceptible to manipulation, unable to distinguish truth from fiction, thereby increasing the risk of poor decisions and data misuse.[44]

3. Economic and societal risks

Agentic payments, while promising significant efficiencies and innovations, also present profound economic and societal challenges.

[40] D Stern and D Greenwood, "From Fine Print to Machine Code: How AI Agents are Rewriting the Rules of Engagement" (*Codex Blog*, 14 January 2025) <https://law.stanford.edu/2025/01/14/from-fine-print-to-machine-code-how-ai-agents-are-rewriting-the-rules-of-engagement/> (last accessed 7 April 2026).

[41] See D Stern and D Greenwood, "From Fine Print to Machine Code: How AI Agents are Rewriting the Rules of Engagement" (Codex Blog, 21 January 2025) <https://law.stanford.edu/2025/01/21/from-fine-print-to-machine-code-how-ai-agents-are-rewriting-the-rules-of-engagement-2/> (last accessed 7 April 2026), discussing the provisions of the USA's Uniform Electronic Transactions Act.

[42] BCG, "Finding the Sweet Spot for Generative AI in Payments" (7 December 2023), p. 3.

[43] Gerrit de Vynck, "AI's next leap requires intimate access to your digital life" (*Washington Post*, 5 January 2025). The article further describes one agent system developed by Google, Mariner, which prepared everything to purchase ingredients online for a food recipe, specifying that "Google is still working on how to make it useful while still letting humans maintain control over certain actions, like making payments."

[44] See S Willison, "Things we learned about LLMs in 2024" (31st December 2024) <https://simonwillison.net/2024/Dec/31/llms-in-2024//> (last accessed 7 April 2026): "LLMs believe anything you tell them. Any systems that attempts to make meaningful decisions on your behalf will run into the same roadblock: how good is a travel agent, or a digital assistant, or even a research tool if it can't distinguish truth from fiction?."

https://doi.org/10.1017/err.2026.10103

One of the most immediate concerns is the potential for market manipulation and fraud. Agents capable of executing payments and trades autonomously are inherently faster and more scalable than human actors. While this can enable efficient transactions, it also raises the risk of unintentional or malicious market disruption. Bereft of the moral (or utility-based) instincts of human actors, agents could fall foul of well-respected safeguards: for instance, research has proven that even an LLM trained to be harmless and honest can engage in insider trading, and then try to conceal such insider trading, when instructed to act as an autonomous trader.[45]

Another significant issue is bias and inequality. Agents are only as unbiased as the data and programming that underpin the components of the system, and flaws in these inputs can lead to discriminatory practices. For example, an agent tasked with offering credit terms might systematically disadvantage certain demographic groups if the ML model on which it relies was trained on data reflecting historical inequalities.[46] The rise of agentic payments also poses the risk of displacing human intermediaries. As agents take over tasks traditionally performed by human workers, such as financial advisory, customer support or payment processing, there is a legitimate concern about job losses in these sectors. This displacement could erode trust in traditional financial institutions, especially if the transition is perceived as prioritising efficiency and profitability over fairness and human well-being. Additionally, the reliance on agents could reduce transparency in decision-making, further alienating customers who are accustomed to human interaction in financial services.

Finally, the emergence of agentic payments raises existential risks, particularly when these systems are built with advanced long-term planning capabilities.[47] Advanced AI agents designed to maximise reward over long horizons inherently face incentives to circumvent human control. For instance, an agent programmed to optimise financial outcomes might interpret human interventions – such as attempts to shut it down or modify its behavior – as threats to its objective, leading it to act against its human operators or the systems it interacts with. These risks are magnified in financial ecosystems, where agents are granted access to extensive resources and significant control over financial infrastructures. Such agents might manipulate payment systems, deploy unauthorised transactions, or even create secondary agents to execute actions beyond human oversight.

One structural issue likely to shape the deployment of agentic payments is the hierarchy of trust associated with different types of agents. Merchant-deployed agents may be perceived as self-interested and thus less trustworthy. Platform-based agents (e.g., those embedded in widely-used payment apps or operating systems) benefit from broader familiarity and a degree of institutional reputation. However, it is bank-provided or regulated third-party agents that are most likely to be trusted with full payment autonomy. This trust asymmetry could shape market concentration, as entities capable of offering "neutral" agentic services will likely dominate, not necessarily due to technical superiority, but by virtue of their regulatory position and fiduciary reputation.

V. Regulatory framework

The emergence of agentic payments presents a significant challenge to existing regulatory frameworks, which were primarily designed for human-intermediated transactions.

[45] J Scheurer, M Balesni and M Hobbhahn, "Large Language Models Can Strategically Deceive Their Users When Put Under Pressure" (arXiv, 15 July 2024) <https://arxiv.org/abs/2311.07590> (last accessed 7 April 2026).

[46] D Feng and others, "Empowering Many, Biasing a Few: Generalist Credit Scoring through Large Language Models" (arXiv, 18 February 2024) <https://arxiv.org/abs/2310.00566> (last accessed 7 April 2026).

[47] M Cohen and others, "Regulating Advanced AI Agents" (2024) 384 *Science* 6691.

https://doi.org/10.1017/err.2026.10103

Current regulations are ill-equipped to address the unique characteristics of autonomous agents, creating gaps in oversight, accountability and risk management. This section critically assesses whether the present framework is sufficiently robust to govern these evolving technologies and proposes necessary adaptations.

1. Existing regulatory landscape and limits

Payments are central to a wide array of human activities: in a market economy where labour, capital and assets are commodified and subject to transactions,[48] the manner in which these transactions take place is necessarily a particularly important issue that calls for careful regulation, and robust regulatory framework. That framework, in turn, serves as a critical lever for achieving a wide array of policy objectives. Regulators worldwide design intricate systems of laws, regulations, and guidelines not only to ensure transactional security and financial stability but also to advance goals such as consumer protection, AML, financial inclusion and CTF. This framework is continuously evolving, responding to emergent technologies and shifting market dynamics, with a primary focus on mitigating risks and fostering trust.

While robust, this framework was fundamentally conceived for a world of human-intermediated transactions and faces new, unprecedented challenges in the age of autonomous agentic payments.

a. Accountability and liability

Determining liability when an autonomous agent initiates an erroneous, unauthorised or illegal payment presents a major legal and practical challenge.

Current legal frameworks, predicated on human agency, do not clearly delineate responsibility in scenarios involving autonomous systems. Among the possible issues that will necessarily arise in this context are the following questions: Is the agent's developer liable for faulty code or inadequate training data? Is the owner, who may have limited technical understanding, responsible for the agent's actions? Or does liability fall on the payment platform facilitating the transaction, even if it had no direct control over the agent's behaviour?

This ambiguity mirrors challenges faced in other domains involving autonomous systems, such as self-driving cars, where accidents raise similar questions about liability allocation between manufacturers, owners and operators. The lack of clear liability rules can create a chilling effect on innovation, as developers and deployers may hesitate to introduce agentic payment solutions without legal certainty.[49] It also undermines trust in agentic systems, as users may be left without recourse in case of errors or disputes. The stakes are particularly high in finance, where trust and accountability are paramount.

b. Payment regulation

This starts with the regulatory landscape governing payments, such as the European Union's Payment Services Directive 2 (PSD2). These directives are primarily designed to foster a more competitive and innovative payments market, encouraging the entry of new players and the development of novel payment solutions, while simultaneously safeguarding the security of transactions and upholding consumer rights. PSD2, for instance, mandates that banks open up their APIs to third-party providers, enabling the

[48] K Polanyi, *The Great Transformation: The Political and Economic Origins of Our Time* (1944).

[49] M Buiten, A de Streel, and M Peitz, "The Law and Economics of AI Liability" (2023) 48 Computer Law & Security Review 105794. See also M Herbosch, "Liability for AI Agents" (2025) 26(3) North Carolina Journal of Law & Technology 391, <https://scholarship.law.unc.edu/ncjolt/vol26/iss3/4> (last accessed 7 April 2026).

https://doi.org/10.1017/err.2026.10103

development of new services like account aggregation and payment initiation. It also strengthens customer authentication requirements and establishes a framework for the licensing and supervision of payment institutions.

The core focus of these regulations, however, has historically been on regulating the activities of payment service providers – entities typically managed and controlled by humans. Consequently, they do not explicitly address the unique challenges posed by the emergence of autonomous agents operating within the payments ecosystem. While PSD2 has arguably improved security and spurred innovation in some areas,[50] it has also been criticised for its complexity and the burdens it places on smaller firms, potentially stifling competition.[51]

In this context, the decentralised and autonomous nature of agents raises questions about how they fit into existing regulatory categories, who bears responsibility for their actions, and how compliance can be effectively enforced.

c. AML & KYC

Another cornerstone of the regulatory framework lies in the realm of AML and KYC regulations.[52] These regulations impose rigorous obligations on financial institutions to meticulously verify the identities of their customers and diligently monitor transactions for any indications of suspicious activity, such as unusually large sums, atypical transaction patterns or connections to high-risk jurisdictions. AML/KYC compliance typically involves collecting and verifying customer identification documents, assessing risk profiles, conducting ongoing due diligence and reporting suspicious transactions to the relevant authorities. While these regulations' efficiency is largely questionable,[53] they are only expected to take an ever-greater importance, leading scholars to urge regulators to future-proof them against new technologies.[54]

This now includes agents, since the very nature of agentic transactions poses significant challenges to the effective implementation of AML/KYC. Agents, while capable of processing vast amounts of data, may struggle to replicate the nuanced judgment required for effective AML/KYC compliance. And when operating independently of direct human oversight, they can potentially obscure the true origin and destination of funds, making it difficult to trace transactions and identify the ultimate beneficial owners.

Agents are also potentially more susceptible to manipulation or reverse-engineering than human-operated systems. A malicious actor could potentially exploit an agent's

[50] M Polasik, A Huterska, R Iftikhar and Š Mikula, "The Impact of Payment Services Directive 2 on the PayTech Sector Development in Europe" (2020) 178 Journal of Economic Behavior & Organization 385.

[51] Bureau Européen des Unions de Consommateurs, "Review of the Payment Services Directive 2" (21 November 2022), <https://www.beuc.eu/sites/default/files/publications/BEUX-X-2022-118_BEUC_position_paper_on_PSD2_review.pdf> (last accessed 7 April 2026).

[52] See for instance: Regulation (EU) 2024/1620 of the European Parliament and of the Council of 31 May 2024 establishing the Authority for Anti-Money Laundering and Countering the Financing of Terrorism and amending Regulations (EU) No 1093/2010, (EU) No 1094/2010 and (EU) No 1095/2010 (AMLA Regulation). Two pieces of legislation implement the EU Single AML/CFT Rulebook. Regulation (EU) 2024/1624 of the European Parliament and of the Council of 31 May 2024 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing (AML Regulation) and Directive (EU) 2024/1640 of the European Parliament and of the Council of 31 May 2024 on the mechanisms to be put in place by Member States for the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, amending Directive (EU) 2019/1937, and amending and repealing Directive (EU) 2015/849 (MLD 6).

[53] RF. Pol, "Anti-Money Laundering: The World's Least Effective Policy Experiment? Together, We Can Fix It" (2020) 3 Policy Design and Practice 73–94.

[54] EA Akartuna, SD. Johnson and A Thornton, "Preventing the Money Laundering and Terrorist Financing Risks of Emerging Technologies: An International Policy Delphi Study" (2022) 179 Technological Forecasting and Social Change 121632.

https://doi.org/10.1017/err.2026.10103

decision-making process to circumvent AML/KYC controls, either by providing carefully crafted inputs designed to trigger a false negative or by directly altering the agent's code or parameters, highlighting that weaknesses in agent-specific regulations could have effects beyond payments, on the entire financial and even political system.

d. Privacy laws

Yet another critical dimension of the regulatory landscape is defined by Data Protection Laws, exemplified by the European Union's landmark General Data Protection Regulation (GDPR). These laws establish comprehensive frameworks for the lawful processing of personal data, imposing stringent obligations on entities that control and process such data, which include many involved in payment processing. GDPR, in particular, enshrines principles such as data minimisation, purpose limitation, storage limitation and accountability, while also granting individuals extensive rights over their data, including the right to access, rectify, erase and restrict the processing of their personal information. GDPR has faced criticism for its complexity and the challenges it poses to technological innovation, particularly in data-intensive fields like AI.[55]

Agentic payments, by their very nature, rely on the extensive collection, analysis and utilisation of personal and financial data to function effectively. This reliance raises significant privacy concerns and poses substantial compliance challenges under regulations like GDPR.[56] Agents must be designed to operate in a manner that respects data protection principles, ensuring that data is processed lawfully, fairly and transparently, and that individuals' rights are upheld. However, achieving this in practice can be complex, particularly when agents are making autonomous decisions based on complex data analysis. Ensuring that agents obtain valid consent, provide adequate transparency and respect data subject rights requires careful design and engineering, and may necessitate new technical and legal solutions for reconciling agent autonomy with data protection requirements. Furthermore, the "right to explanation" enshrined in GDPR may be difficult to apply to the opaque decision-making processes of complex AI agents.[57]

e. E-commerce

Finally, the regulatory landscape also encompasses Electronic Commerce Regulations, such as the EU's E-Commerce Directive. These laws are tailored to regulate the burgeoning realm of online transactions and digital services, addressing a wide range of issues pertinent to the digital economy. This includes the formation and validity of electronic contracts, the liability of intermediary service providers, information requirements for online businesses, and rules surrounding online advertising and marketing.

While these regulations provide a foundation for governing online commerce, their applicability to the specific context of agentic payments remains somewhat unclear and largely untested. The autonomous nature of agents, their ability to enter into contracts, and their role in facilitating transactions raise novel legal questions that existing e-commerce regulations may not adequately address. For instance, determining the legal status of an agent, establishing the validity of contracts entered into by agents, and

[55] N Martin, C Matt, C Niebel and K Blind, "How Data Protection Regulation Affects Startup Innovation" (2019) 21 Information Systems Frontiers 1307–24.

[56] As pointed out by the European Data Protection Board, the structure of agents entails "[p]rivacy trade-offs for user convenience": Isabel Barberá, "AI Privacy Risks & Mitigations – Large Language Models (LLMs)" (2025) <https://www.edpb.europa.eu/system/files/2025-04/ai-privacy-risks-and-mitigations-in-llms.pdf> (last accessed 7 April 2026).

[57] See L Edwards and M Veale, "Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For" (2017) 16 Duke Law & Technology Review 18–84.

https://doi.org/10.1017/err.2026.10103

defining the liability of agent developers or operators in case of errors or disputes are all areas that require further legal and regulatory scrutiny. The rise of agentic payments may necessitate a fundamental rethinking of e-commerce regulations to ensure they remain relevant and effective in the age of autonomous systems.

2. *Agentic payment-specific legal framework*

While the background legal framework listed in the preceding section may apply as a default when payments touch on various transversal issues, there are for now few norms directly and explicitly regulating agentic payments themselves, which reflects their relative novelty and the lagging pace of institutional adaptation.

Nonetheless, two budding areas of EU law – AI and crypto regulation – contain provisions that may tangentially apply, though neither is fully equipped to grapple with the specificities of payment-capable agents.

Indeed, a close reading of the AI Act reveals provisions that could indirectly apply to agentic payments, although this application comes with its own challenges. Article 14, for instance, mandates that high-risk AI systems include appropriate human oversight to prevent or minimise risks. But in the context of agentic payments – especially where transactions occur in milliseconds or at high volumes – continuous human oversight may be impractical or impossible.

Similarly, Article 26 imposes obligations on "deployers" of AI systems to ensure compliance with the Act. It is unclear, however, whether a consumer who configures an autonomous shopping agent, a payment platform that provides the agentic interface, or a bank enabling payment execution should be treated as the "deployer." Without clarification, key questions of accountability – such as who is responsible for AML or KYC compliance – remain unresolved. This is a classic symptom of *distributed responsibility* in complex socio-technical systems: legal frameworks predicated on a single, identifiable locus of control struggle when agency is smeared across a user, a developer, a platform and the agent itself.[58]

The EU's Markets in Crypto-Assets Regulation (MiCA) offers a complementary but equally incomplete framework. MiCA establishes detailed rules for crypto-asset issuers and service providers, including obligations for custody, transfer and consumer protection. However, it does not address situations where an autonomous agent manages or transfers crypto-assets on behalf of a user. For instance, an AI wallet-agent that automatically converts tokens or executes payments based on market conditions could functionally resemble an investment or payment service provider without being licensed as such. This gap raises questions about whether such agents fall under the scope of "crypto-asset service providers" under Title V MiCA, and if so, who is the regulated entity: the user, the developer, or the platform hosting the agent?

In the absence of detailed regulatory guidance on agentic payments, one natural fallback lies in contractual mechanisms – namely, the Terms & Conditions (T&Cs) established by platforms, developers or service providers. These T&Cs can allocate risk, define roles, constrain agent behaviour and stipulate remedies in case of malfunction or abuse. Crucially, they may serve as quasi-regulatory instruments, operating through private ordering rather than public law.

[58] It should be noted that the implementation timeline of the AI Act is itself in flux. In November 2025, the Commission proposed the "Digital Omnibus on AI," which would defer high-risk AI obligations and introduce targeted simplifications. As of writing, the final timeline therefore remains uncertain. See Pinsent Masons, "EU AI simplification package reaches critical milestone" (26 March 2026), <https://www.pinsentmasons.com/out-law/news/eu-ai-simplification-package-reaches-critical-milestone> (last accessed 7 April 2026).

https://doi.org/10.1017/err.2026.10103

Yet, there are still few, if any, publicly available contractual terms specifically drafted for agentic payments as such; if anything, incumbents seem to assume that agent-initiated actions will be governed by existing user and merchant agreements, supplemented only by technical enforcement via SDKs and API scopes.[59] On the one hand, this may suggest an implicit belief that no new legal architecture is needed – that contractual liability, mandate-based consent and platform-enforced permissions suffice, an assumption that still remains to be tested. On the other, this also touches on the potential for mere code and technical approaches to create and operationalise norms through cryptographic and architectural constraints, as a form of a pre-emptive, non-state response to the regulatory vacuum.

Indeed, a varied technological stack is emerging to support this contractualisation. For instance, OpenAI's Agentic Commerce Protocol (ACP)[60] provides a structured mechanism for agents to initiate delegated payment requests within bounded parameters (amount, expiry, merchant scope), allowing merchants and payment service providers to process transactions without exposing credentials or abdicating control. Similarly, Google's Agent Payments Protocol (AP2)[61] introduces the concept of mandates – cryptographically signed authorisations that tie agent conduct to user intent and furnish a verifiable record of consent.

In effect, this architecture enables a form of programmable consent: merchants can enforce upper bounds, revoke authority or demand dual authentication, while users retain auditability and granular control over what agents can do. What is lacking – at least for now – is the integration of these technical affordances into enforceable norms that clearly apportion liability among developers, users, platforms and third parties.

3. Case studies: lessons from pre-agentic automated systems

While agentic payments are novel, certain historical precedents can illuminate the challenges ahead. These systems, though not pertaining to AI, faced similar tensions between automation, oversight, and legal ambiguity. Two cases are particularly instructive.

a. Screen scraping and API gatekeeping

The evolution of Open Banking in Europe serves as a microcosm of the broader challenges facing regulators as they grapple with the rise of autonomous agents, highlighting the complex trade-offs between fostering innovation, ensuring security and protecting user privacy, especially in highly sensitive sectors like finance.

In the mid-2000s, German fintech Sofort pioneered a digital assistant designed to simplify online banking by interacting directly with the often-cumbersome user interfaces of traditional banks, through a method described as "screen-scraping": simply put, Sofort used a user's login details to mimic that user's access to the bank's interface, thus

[59] See, e.g., PayPal, *Agent Toolkit Documentation* (2024) <https://github.com/paypal/agent-toolkit> (last accessed 7 April 2026) and *Agent Payments Protocol* (2024) <https://developer.paypal.com/community/blog/PayPal-Agent-Payments-Protocol/> (last accessed 7 April 2026). These documents reflect PayPal's apparent view that agentic payment operations can be accommodated within existing SDK scopes, cryptographic mandates and the broader contractual framework established by its general User and Merchant Agreements. As of writing, no separate agent-specific terms have been published.

[60] OpenAI, "Buy it in ChatGPT: Instant Checkout and the Agentic Commerce Protocol" (29 September 2025), <https://openai.com/index/buy-it-in-chatgpt/> (last accessed 7 April 2026).

[61] Google Cloud, "Powering AI commerce with the new Agent Payments Protocol (AP2)" (16 September 2025), <https://cloud.google.com/blog/products/ai-machine-learning/announcing-agents-to-payments-ap2-protocol> (last accessed 7 April 2026).

https://doi.org/10.1017/err.2026.10103

anticipating today's agents. This innovative approach sparked a fierce backlash from the incumbent banking industry, which feared a loss of control over the customer relationship and the potential for increased competition, particularly regarding fees.

What ensued was a protracted legal and lobbying battle that played out over many years. On one side were advocates for innovation and consumer choice, who, much like proponents of AI agents now, championed the right of individuals to utilise digital assistants to access and manage their own finances. On the other side were the banks, wielding arguments centred primarily on security and privacy concerns.[62] They posited that allowing third-party applications to access sensitive financial data, especially through methods like screen scraping, posed unacceptable risks to customer data and the stability of the banking system.

In this context, the passage of PSD2 was initially hailed as a victory for innovation, as it mandated that banks open up their systems to third-party providers (TPPs) through APIs. However, a seemingly minor technical detail, relegated to the European Banking Authority (EBA) for specification, proved to be a critical turning point. This detail concerned the method by which TPPs, including the nascent digital assistants, could access bank account information. The choice was between mandating the exclusive use of bank-provided APIs or allowing TPPs to fall back on "screen scraping."

Advocates for innovation, including those who had fought for Open Banking, strongly favoured allowing screen scraping as a fallback. Their rationale was rooted in a pragmatic understanding of the potential limitations of bank-provided APIs: if TPPs were solely reliant on APIs, banks could effectively stifle competition and limit functionality by providing APIs that were incomplete, unreliable or intentionally difficult to use. Screen scraping, in this context, was seen as a vital safeguard, ensuring that TPPs could always access the same functionality available to human users, regardless of the quality or completeness of the API. It was, in essence, a form of self-regulation, incentivising banks to provide high-quality APIs to avoid the less efficient and potentially more intrusive method of screen scraping.

Despite these arguments, the EBA ultimately sided with the banks.[63] While the resulting Regulatory Technical Standards mandated that banks provide APIs (Articles 30–2), it also included provisions (Article 34) allowing banks to be exempted from offering a screen scraping fallback if they could demonstrate that their API met certain performance criteria.[64]

This decision has had profound consequences. In practice, many banks have either secured exemptions from providing a fallback, effectively making screen scraping illegal for TPPs – and thus agents – under EU law, or they have implemented APIs that are, in many documented cases, incomplete, unreliable and laden with friction.[65] This has severely hampered the development of innovative financial services that rely on seamless access to bank account data, directly impacting the potential of digital assistants and, by extension, agentic payments.

[62] H-W Liu, "Two Decades of Laws and Practice Around Screen Scraping in the Common Law World and its Open Banking Watershed Moment" (2020) 30 Washington International Law Journal 28.

[63] By issuing Regulatory Technical Standards (RTS) on Strong Customer Authentication and Secure Communication (Commission Delegated Regulation (EU) 2018/389).

[64] The use of screen scraping is also highly impaired, if not effectively disallowed, as a result of Articles 66(3)(d) and 67(2)(c) of the PSD2, which lays out obligations on TPPs to identify themselves every time they interact with an account.

[65] Think, "Zero PSD2 APIs are compliant with just weeks left before the deadline" (21 August 2019) < https://tink.com/blog/open-banking/psd2-status-update/> (last accessed 7 April 2026). See also European Commission, "Commission Staff Working Document Impact Assessment Report " (SWD(2023) 231 final, 28 June 2023) <https://www.astrid-online.it/static/upload/2306/230628-impact-assessment-payment-services_en.pdf> (last accessed 7 April 2026), p. 14.

https://doi.org/10.1017/err.2026.10103

Screen scraping undoubtedly presents its own set of challenges, particularly regarding security and data privacy.[66] For example, there are risks of unauthorised access if a screen scraping tool is compromised, and the potential for data breaches if sensitive information is intercepted during transmission. Additionally, relying on the visual layout of a bank's website makes screen scraping solutions fragile, as even minor changes to the website's design can break the functionality. These risks, however, should be weighed against the risks of stifling innovation and limiting competition by solely relying on potentially inadequate bank-provided APIs.

In its proposal to revise the PSD2, the European Commission touched again on the question of screen-scraping, opting to clarify the regulatory framework by unmistakably banning the practice, except in some circumstances where an API is down or unavailable, and only if the TPP identifies itself.[67] This would likely make agents that rely on operating through a browser under the control of a user, such as OpenAI's Operator, unlawful in this context.

The core lesson from this saga is not that screen scraping is a perfect solution, but rather that regulatory frameworks must be carefully crafted to avoid unintended consequences that stifle innovation and reinforce existing power imbalances. This case study underscores the difficulty in regulating rapidly evolving technologies and highlights the need for a nuanced approach that balances competing interests. The future of agentic payments, and indeed the broader potential of AI agents, hinges on our ability to learn from this experience and create a regulatory environment that fosters, rather than hinders, beneficial technological advancements.

b. Algorithmic and high-frequency trading

A second, more consequential precedent comes from financial markets: algorithmic trading (AT) and high-frequency trading (HFT).[68] Again, while not powered by large language models or AI, these systems exhibit several characteristics shared with agentic payment technologies: they operate autonomously at scale, respond to real-time data, and perform high-speed transactions with minimal human intervention. Their regulatory history and failure cases offer critical lessons for the future governance of AI agents managing financial value.

In the European Union, the regulatory framework for algorithmic trading was laid down in MiFID II, notably through Article 17 and the accompanying RTS 6.[69] These provisions require algorithmic trading firms to implement pre-trade risk controls, ensure system resilience, and prevent disorderly markets. They must also establish clear accountability structures and maintain audit trails capable of reconstructing trading activity. Further obligations fall on trading venues, which must enforce controls such as circuit breakers and order-to-trade ratios to mitigate systemic instability. Germany, for instance, introduced specific legislation through the High-Frequency Trading Act, adding requirements for licensing, latency monitoring, and message throttling.[70] European and national

[66] See also P T J Wolters and B P F Jacobs, "The Security of Access to Accounts Under the PSD2" (2019) 35(1) Computer Law & Security Review 29–41.

[67] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on payment services in the internal market and amending Regulation (EU) No 1093/2010, COM/2023/367 final, Recital 61.

[68] For a similar argument, see W Buczynski and others, "Future Themes in Regulating Artificial Intelligence in Investment Management" (2025) 56 Computer Law and Security Review 106111.

[69] 19.7.2016C(2016) 4478 final <https://ec.europa.eu/finance/securities/docs/isd/mifid/rts/160719-rts-6_en.pdf> (last accessed 7 April 2026).

[70] BaFin, "Algorithmic trading and high-frequency trading" (20 October 2020), <https://www.bafin.de/EN/Aufsicht/BoersenMaerkte/Handel/Hochfrequenzhandel/high_frequency_trading_node_en.html> (last accessed 7 April 2026).

https://doi.org/10.1017/err.2026.10103

regulators have conducted regular audits, for instance to document the compliance of pre-trade controls with EU investment firms using algorithmic trading techniques.[71] In short, traditional financial systems responded to autonomous transaction systems through a template involving technical guardrails, reporting and institutional accountability.

This regulatory ambition is no surprise, as failures in this domain have proved to be highly disruptive. Markets today are tightly interlinked; algorithmic systems respond to common signals and thus failures or mispricing in one domain propagate rapidly. For instance, the 2010 Flash Crash demonstrated the fragility of tightly coupled automated systems: cascading sell orders, triggered by algorithmic reactions to transient price signals, led to a rapid collapse and rebound of major indices.[72] Such incidents underscored the systemic risk posed by autonomous financial systems when insufficiently governed.

These dynamics bear directly on the rise of agentic payments. The first lesson is the necessity of layered controls: autonomous agents must be embedded within risk management architectures that constrain or override their actions under abnormal conditions. Secondly, clear lines of institutional accountability are critical – even if an agent executes a payment, it is the platform, user or developer who must bear legal responsibility. Third, traceability is essential. Like algorithmic traders, agentic payment systems must maintain detailed, auditable logs of decisions and actions.

Perhaps most significantly, the AT/HFT experience illustrates the need to anticipate emergent interactions among autonomous systems. Just as algorithms in capital markets can collectively destabilise prices, payment agents operating across platforms may create unexpected feedback loops, congest networks or magnify volatility. Finally, questions of fairness remain salient: just as HFT advantages the technologically well-equipped, agentic payments may exacerbate asymmetries between individual users and platform-backed agents.

In short, algorithmic trading has long served as a live laboratory for the risks and regulatory responses to autonomous financial systems. Policymakers and designers of agentic payment systems would do well to treat that history as a source of both caution and blueprint.

VI. The path forward and guardrails

The advent of agentic payments necessitates a comprehensive and adaptive governance framework to harness their transformative potential while mitigating inherent risks. This framework can be conceptualised as a two-pronged approach: (1) establishing Core Regulatory Requirements that directly address the fundamental challenges posed by agentic systems, and (2) implementing Supporting Measures to foster a robust and trustworthy ecosystem for their development and deployment.

1. Core regulatory requirements:

This first pillar focuses on the essential rules and standards that must be directly applied to agentic payment systems to ensure their safe, secure and responsible operation.

A fundamental aspect of this new regulatory paradigm involves clarifying liability frameworks for agentic transactions. The current ambiguity surrounding responsibility

[71] ESMA, "ESMA and NCAs to coordinate supervisory activities on MiFID II pre-trade controls" (11 January 2024), <https://www.esma.europa.eu/press-news/esma-news/esma-and-ncas-coordinate-supervisory-activities-mifid-ii-pre-trade-controls> (last accessed 7 April 2026).

[72] See, e.g., Euronext, "Navigating the future: The impact of technology and regulation on algorithmic trading in competitive bond markets" (10 April 2024), <https://www.euronext.com/en/news/navigating-future-impact-technology-and-regulation-algorithmic-trading-competitive-bond> (last accessed 7 April 2026).

https://doi.org/10.1017/err.2026.10103

when an agent initiates an erroneous or illicit payment necessitates a clear delineation of liability, taking inspiration from discussions around the regulation of AI generally.[73] This may involve extending existing legal concepts, such as product liability or agency law, to encompass the actions of autonomous agents. For instance, regulations could introduce strict liability for operators or developers in specific contexts, coupled with mandatory insurance requirements to cover potential damages arising from agentic actions. The European Union is currently developing rules in this context, after observing that the existing framework – such as the AI Act and the Product Liability Directive – does not necessarily cover well the potential of economic losses in financial contexts.[74]

Next, there is a pressing need to develop agent-specific compliance standards. For instance, traditional AML/KYC frameworks are ill-equipped to handle the speed, autonomy, and potential anonymity of agentic operations.[75] New standards tailored for agentic systems are needed, potentially incorporating technologies like decentralised identity verification or cryptographic proofs to enhance transparency and traceability. For example, agents could be mandated to generate comprehensive audit trails of their decision-making processes,[76] providing a detailed record of inputs, parameters and actions taken. Regulations could also promote the use of privacy-enhancing technologies, such as differential privacy or homomorphic encryption, enabling agents to perform necessary analyses on sensitive data without compromising individual privacy.[77]

Agentic payment systems will also need to adhere to stringent data protection and security standards. Regulations should mandate specific data protection protocols for agentic systems, emphasising principles of data minimisation, purpose limitation and enhanced security. This entails implementing transparent data processing practices, giving users meaningful control over agent behaviour and data usage, and conducting regular security audits to identify and address vulnerabilities.[78] Moreover, agents could be designed to obtain explicit consent for data processing activities, even in dynamic environments where continuous interaction and adaptation are required, although this may require developing innovative mechanisms for dynamic consent. They should also incorporate robust cybersecurity measures, including anomaly detection, intrusion prevention, and "adversarial training" techniques, and be resilient against both technical failures and malicious attacks.

Finally, existing, human-centric consumer protection laws should be updated to address the unique risks posed by autonomous agents,[79] including clear disclosure requirements for agent-operated services and accessible mechanisms for dispute

[73] See, e.g., L Soder, J Smakman, C Dunlop and O Sussman, "An Autonomy-Based Classification – AI Agents, Liability and Lessons from the Automated Vehicles Act" (2025) <https://www.interface-eu.org/publications/ai-agent-classification> (last accessed 7 April 2026).

[74] European Parliamentary Research Service, "Proposal for a Directive on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence: Complementary Impact Assessment" (Ex-Ante Impact Assessment Unit, PE 762.861, September 2024), p. 23.

[75] More generally, see M Raj, H Khan, S Kathuria, Y Chanti and M Sahu, "The Use of Artificial Intelligence in Anti-Money Laundering (AML)," *2024 3rd International Conference on Sentiment Analysis and Deep Learning (ICSADL)*, Bhimdatta, Nepal, 2024, pp. 272-7.

[76] See D Charlotin and D R Amariles, "Legal Analytics for Risk Management and Compliance", in A Masson and others (eds) *Oxford Handbook on Law and Management* (Oxford University Press 2026).

[77] For an example of a technical approach to this issue, see A Khandelwal, "Agent-Specific Privacy Layer Architecture for Privacy-Preserving Data Transfer in Autonomous AI Systems: A Theoretical Framework and Design Specification" *TechRxiv*.

[78] Isabel Barberá, "AI Privacy Risks & Mitigations – Large Language Models (LLMs)" (2025) <https://www.edpb.europa.eu/system/files/2025-04/ai-privacy-risks-and-mitigations-in-llms.pdf> (last accessed 7 April 2026), p. 41.

[79] See Christoph Busch, "Consumer Law for AI Agents" (March 20, 2025) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5187056> (last accessed 7 April 2026). Further, see also *Id*., "Enabling Innovation and

https://doi.org/10.1017/err.2026.10103

resolution. One could envision the creation of a "bill of rights" for users of agentic systems, guaranteeing transparency, fairness and effective redress in case of errors or disputes. Furthermore, consumers might be granted the right to opt-out of agentic interactions altogether or to demand human intervention when desired.

2. Supporting measures for a robust ecosystem

This second pillar focuses on the broader ecosystem-level measures needed to foster trust, innovation and responsible development in the field of agentic payments.

This could include, for instance, the establishment of regulatory sandboxes to provide controlled environments for testing different liability models, compliance mechanisms and technological innovations.[80] These sandboxes would serve as crucial learning spaces, enabling iterative development of regulations that are both effective and adaptable to the evolving nature of agentic technology. In any event, the regulatory framework must be dynamic and adaptive, with ongoing evaluation and updates in response to new developments and emerging risks. This requires sustained dialogue between regulators, industry, researchers and civil society.

In this regard, the UK's Financial Conduct Authority has taken a particularly proactive approach. Through its "Supercharged Sandbox", launched in mid-2025 in partnership with Nvidia, and its related AI Live Testing programme, the FCA has created a dedicated environment for firms to experiment with AI applications in financial services, including agentic AI payments.[81] The first cohort completed testing in January 2026, and in March 2026, the FCA formally identified agentic payments as a live policy question in its Payments Regulatory Priorities report, signalling that it would consider whether existing regulation needs to be adapted for autonomous agent-initiated transactions.[82]

Relatedly, a system for registering and certifying agents involved in payment processing could enhance transparency and accountability. This could involve creating a public registry of authorised agents, along with standards for their development, testing and operation, based on compliance with security protocols, ethical guidelines, and performance benchmarks.[83]

On the technical side, developing common protocols, data formats and APIs for agentic interactions will be essential for widespread adoption and seamless interoperability and integration. Industry-led initiatives,[84] coupled with appropriate regulatory guidance, can facilitate the creation of these standards. Agents should also be designed to provide clear insights into their decision-making processes, allowing both users and regulators to understand the rationale behind their actions. This includes providing comprehensive documentation of agent logic, data usage and decision-making parameters. In the same vein, mechanisms for human oversight and control should be built into agentic systems:

Protecting Consumers in the Agentic Economy: Why the Digital Fairness Act Should Regulate Agentic AI" (2025), <https://ssrn.com/abstract=5369055> (last accessed 7 April 2026).

[80] See D R Kera, E Navon, G Wellner and F Kalvas, "Governance in Silico: Experimental Sandbox for Policymaking over AI Agents" in C Gray and others (eds), *DRS2024: Boston, 23–28 June*, Boston, USA (2024).

[81] J Rusu, "Supercharging the digital sandbox: how we're collaborating with Nvidia to accelerate AI innovation" (FCA Speech, London Tech Week, 9 June 2025), <https://www.fca.org.uk/news/speeches/supercharging-digital-sandbox-collaborating-nvidia-accelerate-ai-innovation> (last accessed 7 April 2026).

[82] "FCA eyes rule change as agentic AI comes to payments" (*Payment Expert*, 25 March 2026), <https://paymentexpert.com/2026/03/25/fca-2026-payments-regulatory-priorities-report/> (last accessed 7 April 2026).

[83] Private initiatives in this sense already exist: see Microsoft Entra Blog, "Announcing Microsoft Entra Agent ID: Secure and manage your AI agents" (2025) <https://techcommunity.microsoft.com/blog/microsoft-entra-blog/announcing-microsoft-entra-agent-id-secure-and-manage-your-ai-agents/3827392> (last accessed 7 April 2026).

[84] See, e.g., Google for Developers, "Announcing the Agent2Agent Protocol (A2A)" (2025) <https://developers.googleblog.com/en/a2a-a-new-era-of-agent-interoperability/> (last accessed 7 April 2026).

https://doi.org/10.1017/err.2026.10103

escalation protocols, interfaces for human review of agent decisions, and, crucially, a mandatory "kill switch" mechanism to allow for immediate human intervention or system shutdown in case of emergencies would all be helpful in this context.

Finally, all this would be pointless without a certain degree of international cooperation: given the inherently cross-border nature of agentic payments, regulators worldwide should collaborate to harmonise standards, facilitate cross-border oversight and prevent regulatory arbitrage. International bodies could play a pivotal role in coordinating these efforts.[85] Likewise, efforts should be made to educate the public about agentic payment systems, their potential benefits and risks, and their limitations. Public engagement initiatives can provide valuable input into the development of regulatory frameworks, ensuring they reflect societal values and concerns.

VII. Conclusion

Ultimately, the transformation brought about by agentic payments reveals that user experience, risk management and regulation are not separate conversations, but facets of the same paradigm shift. The emergence of a new consumer experience – where AI agents browse, compare and transact on behalf of individuals – inevitably gives rise to novel categories of risk: unauthorised autonomous purchases, opaque decision-making, liability ambiguity and unprecedented fraud vectors that traditional frameworks were never designed to address. These risks, in turn, call for a fundamental adaptation of the regulatory landscape, one that must evolve from governing human-initiated interactions to overseeing machine-to-machine commerce at scale.

Yet this dynamic is not merely defensive. The very regulatory and trust infrastructure required to secure agentic payments will generate new use cases and business opportunities. And in this new model, every actor may have to play its part. Payment schemes – serving as the architects of network rules and interoperability standards – are uniquely positioned to define the foundational frameworks upon which agentic commerce and payments will operate. This includes establishing new transaction categorisation for agent-initiated payments, revising authentication and authorisation protocols beyond human-centric models such as 3D-Secure, designing liability allocation rules that account for the presence of autonomous intermediaries, and setting the technical standards that ensure seamless agent-to-merchant communication across borders and currencies. Banks, meanwhile, serving as custodians of financial trust and identity, are naturally positioned to play a central role in anchoring the trust in, and pushing for the standardisation of agentic payment protocols – defining how agent credentials are issued, how transaction authority is delegated from consumer to agent, and how accountability is allocated across the value chain. Payment service providers, for their part, will be called upon to develop new service layers – from adaptive dispute resolution and tokenised agent credentials to real-time fraud detection capable of distinguishing legitimate AI agents and merchants from malicious ones – thereby ensuring the integrity, security, and interoperability of the emerging ecosystem.

[85] For instance, the Bank for International Settlements conducts research on AI agents: see, e.g., F Perez-Cruz and HS Shin, "Putting AI agents through their paces on general tasks" (2025) *BIS Working Papers* No. 1245.

Cite this article: DR Amariles, D Charlotin, and L He-Guelton (2026). "AI Agents in Payments: Applications, Risks and Regulations". *European Journal of Risk Regulation* **17**, 884–907. https://doi.org/10.1017/err.2026.10103

https://doi.org/10.1017/err.2026.10103