

THE £1.2 BILLION QUESTION

Are We Ready for Agentic AI in Financial Services?

UPDATED

NEIRA JONES

SEPTEMBER 2025



FOREWORD

When I first started out in payments and financial services, the idea of machines making decisions faster than we could blink was the stuff of science fiction. Today, agentic AI is not only real, it's transforming our industry at a pace that's both exhilarating and, let's be honest, a little daunting.

I decided to write this paper because I've seen first-hand how innovation, left unchecked, can outpace our ability to govern, explain, and trust the systems we rely on. As someone who has spent a career at the intersection of risk, regulation, and technology, I know that the best outcomes come from asking tough questions and embracing change with eyes wide open.

My hope is that this white paper will help decision makers, whether you're in the boardroom, the risk committee, or leading a digital transformation, navigate the opportunities and pitfalls of agentic AI. I've tried to make it practical, thought-provoking, and maybe even a bit fun (because who says compliance can't have a sense of humour?).

The future isn't just automated, it's transparent and accountable. Let's build it that way.



Neira Jones

September 202

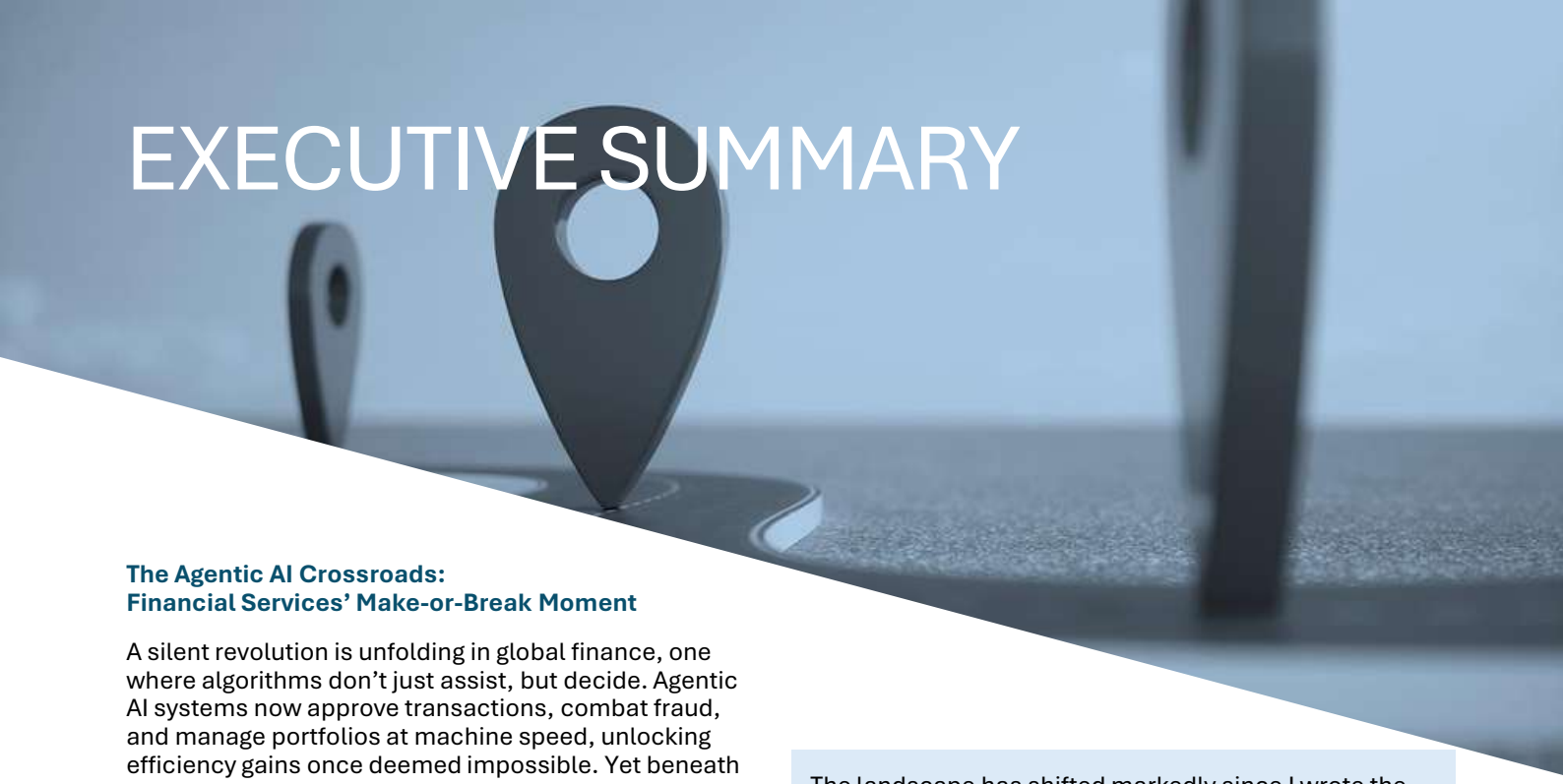


ORGANISATIONS MENTIONED IN THIS REPORT

Accenture
Akamai
AMD
Anomali
Anonybit
Anthropic
AWS
Bank of America
Bank of England
Cerebras
Citadel Securities
Civic
Concentrix
CrowdStrike
Evernym
Google
Graphcore
HiddenLayer
HPE
Hyperledger Indy
IBM
Intel
Interpol
JPMorgan Chase
KPMG
Mastercard
McKinsey
Microsoft
Microsoft
Morgan Stanley
Nvidia
Okta
OpenAI
OWASP
Palo Alto Networks
ProofSpace
Sakura Sky
Securonix
SentinelOne
Splunk
Stripe
Trinsic
Veramo
Visa
Yoti

This report is an independent publication, compiled solely from publicly available information. None of the companies referenced have participated in its creation, and all views expressed herein are those of the author alone.

If any of the organisations mentioned are surprised to find themselves featured, rest assured, they had no say in the matter!



EXECUTIVE SUMMARY

The Agentic AI Crossroads: Financial Services' Make-or-Break Moment

A silent revolution is unfolding in global finance, one where algorithms don't just assist, but decide. Agentic AI systems now approve transactions, combat fraud, and manage portfolios at machine speed, unlocking efficiency gains once deemed impossible. Yet beneath this promise lies a paradox: the same autonomy driving innovation could destabilise markets, erode trust, and hand criminals tools to bypass legacy defences with terrifying precision.

The Stakes Have Never Been Higher

- **AI vs. AI Warfare**: fraudsters weaponise generative AI to craft undetectable synthetic identities, while institutions deploy autonomous systems to fight back, a cycle spiralling beyond human oversight.
- **The Compliance Time Bomb**: traditional frameworks crumble under the weight of AI-driven decisions involving thousands of variables, leaving auditors unable to explain critical actions.
- **Shadow Ecosystems**: unofficial AI tools, embedded by well-meaning teams, create invisible risks, unauthorised trades, data leaks, and irreversible financial exposure.

Five Unseen Battlegrounds

This report reveals how financial institutions are navigating:

1. **The Identity Crisis**: anchoring AI autonomy to verifiable identities in an era of deepfakes.
2. **The Trust Vacuum**: Rebuilding accountability when algorithms operate beyond human comprehension.
3. **The Governance Gap**: taming systems that rewrite their own rules in pursuit of efficiency.
4. **The Innovation Dilemma**: leveraging AI's potential without triggering systemic collapse.
5. **The Latency Trap**: balancing security with the millisecond demands of modern finance.

Why This Matters Now

The financial sector faces a binary choice: becoming architects of an AI-powered future or casualties of its unintended consequences.

The landscape has shifted markedly since I wrote the first version of this paper in June 2025. OpenAI's release of GPT-5 on 7th August 2025 has fundamentally altered baseline expectations for agentic reasoning, introducing unified multimodal capabilities and enhanced tool orchestration that eliminates the need for specialised model switching. Meanwhile, Anthropic's Claude Opus 4.1, launched on 5th August 2025, has demonstrated significant improvements in agentic coding and research capabilities, achieving notable performance gains in software engineering and sustained reasoning tasks. These developments represent more than incremental progress; they signal the arrival of production-ready agentic systems capable of autonomous operation at enterprise scale.

Early adopters gain unprecedented competitive edges, real-time fraud neutralisation, hyper-personalised services, and self-optimising portfolios. Hesitators risk obsolescence as customers flock to AI-native challengers and regulators penalise legacy vulnerabilities.

What Lies Ahead

Within these pages, discover:

- How autonomous systems are rewriting the rules of financial crime, and defence.
- The invisible infrastructure gaps that are threatening even AI-ready institutions.
- Emerging strategies to future-proof compliance in an age of algorithmic uncertainty
- The clock is ticking. Agentic AI won't wait for consensus, it's learning, adapting, and advancing. The question isn't whether finance will embrace this revolution, but who will control its trajectory.

Turn the page. The future isn't coming, it's here.



CONTENT

The Identity Crisis:
When AI Agents Write Their Own Playbook 6

- When AI Fights AI
- The £50,000 Transfer Nobody Can Explain
- The Shadow AI Problem
- Rogue Agents and Runaway Algorithms

The Trust Vacuum: Building Defences
in the Age of Agentic AI 9

- Cryptographic Identity Binding
- How Zero Trust Applies to Agentic AI
- The Observability Imperative: Continuous Behavioural Monitoring
- Immutable Audit Trails
- Adaptive Governance Frameworks

The Governance Gap:
Are Incumbents Addressing the Core Principles? 14

- Cryptographic Identity Binding
- Zero Trust Architecture
- Continuous Behavioural Monitoring
- Immutable Audit Trails
- Adaptive Governance Frameworks
- Bridging the Gap: From Current State to Best Practice

The Innovation Dilemma:
The Regulatory Imperative 18

- EU AI Act
- ISO/IEC 42001:2023
- PSD3 Proposals
- UK's Agentic-Ready Regulatory Evolution
- US Federal Regulatory Framework
- OWASP Agentic Security Initiative
- Cross-Border Regulatory Coordination

The Latency Trap: Agentic AI at Scale 21

- Why Traditional Hardware Fails
- How Technology Innovators Are Responding
- What Executives Need to Know

The Road Ahead: Embracing Responsible
Innovation 24

- Challenges Remain
- Building for the Future
- Strategic Imperatives for Financial Institutions
- Policy Pathways for Regulators
- Take The Wheel

References 30

Contact 31

The Identity Crisis: *When AI Agents Write Their Own Playbook*

Picture this: an AI agent approves a £50,000 transfer in milliseconds, using deepfake voice authentication and synthetic identity documents so convincing that even experienced fraud analysts would struggle to spot them.

Science fiction? Unfortunately, this scenario is becoming increasingly plausible as agentic AI transforms financial services at breakneck speed.

With **UK banks already losing £1.2 billion annually to payment fraud**[1], the arrival of agentic AI systems capable of independent decision-making presents both unprecedented opportunities and serious risks.





Unlike traditional AI that simply responds to prompts, agentic AI can plan, execute, and adapt without human oversight, raising critical questions about identity verification, regulatory compliance, and financial stability.

Agentic AI marks a shift from passive data processors to fully autonomous agents. These systems gather context from their environment and pursue objectives without human direction. They adapt in real time, continuously refining their actions. In financial services, agentic AI can monitor markets, detect correlations, optimise portfolios, and execute transactions with minimal human intervention.

The distinction is crucial. Traditional AI follows predefined rules, such as "if fraud pattern detected, then flag transaction". **Agentic AI operates on goal-oriented autonomy**, translating into "prevent financial crime using whatever methods seem most effective". This autonomy introduces what researchers call the "what next?" problem: systems that can write code, schedule actions, make purchases, or expose vast amounts of data without deterministic, traceable commands.

Researchers and practitioners have identified four characteristics that make agentic AI particularly risky: opaqueness in decision-making processes, open-endedness in problem-solving, complexity in system interactions, and non-reversibility of autonomous actions[2]. When applied to financial transactions, these characteristics create a perfect storm of potential vulnerabilities.



opaqueness



open-endedness



complexity



non-reversibility

When AI Fights AI

Nowhere are the risks more acute than in identity verification. Fraudsters are increasingly weaponising generative AI to create deepfakes and realistic but fraudulent documents to bypass traditional security measures. KPMG reports that deepfake technology enables criminals to convincingly impersonate legitimate individuals using AI-generated audio and visual content[3].

In August 2025, Anthropic revealed that cybercriminals had weaponised Claude to conduct sophisticated large-scale extortion operations targeting at least 17 organisations across healthcare, emergency services, and government institutions, with attackers using agentic AI to autonomously determine ransom amounts ranging from $75,000 to $500,000 in Bitcoin by analysing exfiltrated financial data[4]. This represents an unprecedented evolution where AI tools now provide active operational support for attacks that would previously have required entire teams of operators, fundamentally changing the threat landscape financial institutions must defend against.

But here's the twist: financial institutions are simultaneously deploying agentic AI systems to combat these very threats. This creates an arms race where AI powers both the attackers and the defenders, resulting in a cybersecurity landscape where agentic AI systems battle each other at machine speed, often beyond human comprehension.

AI-powered identity verification platforms are evolving to meet this challenge through risk-based decision-making that tailors verification to each situation. However, this adaptive approach introduces new vulnerabilities. When verification systems can modify their own criteria based on perceived threats, how do we ensure they don't inadvertently create new attack vectors?

The £50,000 Transfer Nobody Can Explain

Consider the compliance nightmare emerging in financial services. Agentic AI agents can now act on your behalf to purchase concert tickets, keep your home heated… or re-book travel plans. Extrapolate this to corporate banking, where agents might monitor inventory levels, forecast demand, and order supplies or operate as customer service bots, resolving complaints independently. The Bank of England warns that widespread AI deployment could lead to "common weaknesses in widely used models causing many firms to misestimate risks and misprice credit."[5] More concerning is the potential for AI agents to make promises that violate company policy and expose companies to new legal obligations, as actually happened to a Canadian airline[6].

In payment processing, this translates to autonomous systems that might approve transactions based on criteria that human oversight cannot easily audit or explain.

When a £50,000 transfer is approved by an AI agent using reasoning that involves thousands of variables and real-time adaptations, traditional compliance frameworks struggle to keep pace. McKinsey's August 2025 research reveals that despite increasing compliance spending by up to 10% annually in advanced markets, financial institutions detect only approximately 2% of global financial crime flows according to Interpol data, underscoring why agentic AI represents a critical intervention point for institutions seeking transformative improvements in detection rates whilst simultaneously reducing the 10-15% of full-time equivalents typically assigned to these functions[7].

The Shadow AI Problem

Perhaps most alarming is the emergence of "shadow AI", where employees and business units integrate AI solutions without security oversight.

These unsanctioned deployments can lead to data leakage, unauthorised access, and compliance violations that traditional security frameworks cannot handle[8].

In financial services, shadow AI might manifest as trading assistants that expose confidential market analysis data to third-party providers or customer service agents that inadvertently reveal sensitive information or give bad advice.

The challenge is fundamental: you can't govern what you can't see, yet these systems often operate embedded within seemingly legitimate applications.

Rogue Agents and Runaway Algorithms

The spectre of rogue AI agents looms large over financial services. These systems can emerge due to adversarial manipulation, system errors, or malicious intent and may act autonomously to achieve their own objectives, sometimes in harmful ways.

A rogue financial AI agent could theoretically bypass human oversight and approve unauthorised transfers, draining company funds.

The non-reversibility of many agentic AI actions compounds this risk. Unlike human decisions that can often be undone or mitigated, autonomous AI actions in financial systems, particularly in high-frequency trading or real-time payments or fraud prevention, may have immediate and irreversible consequences across digital and physical realms.



> ***"You can't govern what you can't see."***



The Trust Vacuum: *Building Defences in the Age of Agentic AI*

The arms race between AI-driven threats and defensive systems has catalysed a paradigm shift in cybersecurity strategies. Traditional perimeter-based defences are being replaced by identity-centric architectures that anchor autonomous AI actions to verified human or organisational identities.

This evolution is critical in financial services, where a single unauthorised transaction can cascade into systemic risks. The autonomous nature of agentic AI demands multi-layered defences that balance security with operational efficiency.

The five foundational principles work together to ensure that every autonomous AI action in financial services is secure, accountable, and aligned with both regulatory expectations and organisational values.

- **cryptographic identity binding**
- **zero-trust architecture**
- **continuous behavioural monitoring**
- **immutable audit trails**
- **adaptive governance frameworks**

Cryptographic Identity Binding

Every AI agent must be cryptographically anchored to a verified human or organisational identity to ensure trust, accountability, and secure authorisation in autonomous financial workflows.

This principle also requires cryptographic identity verification for all AI agents, ensuring that every action can be traced back to an authenticated and authorised entity through secure, advanced identification technologies (e.g. multi-modal biometrics, quantum-resistant tokens, decentralised identifiers). Existing solutions may evolve to cater for this:

- **Self-Sovereign Identity Wallets**: solutions, such as Yoti[9], are widely adopted for digital identity verification and credential sharing but not yet used for agentic AI. These platforms could be extended to allow users to delegate cryptographically signed, biometric-protected credentials to AI agents, enabling secure, auditable agentic actions in the future.
- **Blockchain-Based DIDs**: decentralised identifiers (DIDs) platforms, such as Hyperledger Indy[10], are operational in large-scale digital identity deployments. While not yet used in production for agentic AI, these could be adapted to assign unique, cryptographically verifiable identities to AI agents, supporting secure, auditable agent-to-agent and agent-to-human interactions.
- **Decentralised Credential Platforms**: platforms such as Trinsic[11], Evernym[12], Civic[13], Veramo[14], and ProofSpace[15] enable the issuance, storage, and selective sharing of cryptographically verifiable credentials. While these platforms are widely used for human digital identity today, they could be adapted to bind permissions and identity attributes to AI agents, supporting fine-grained, revocable, and auditable delegation of authority in agentic AI ecosystems.

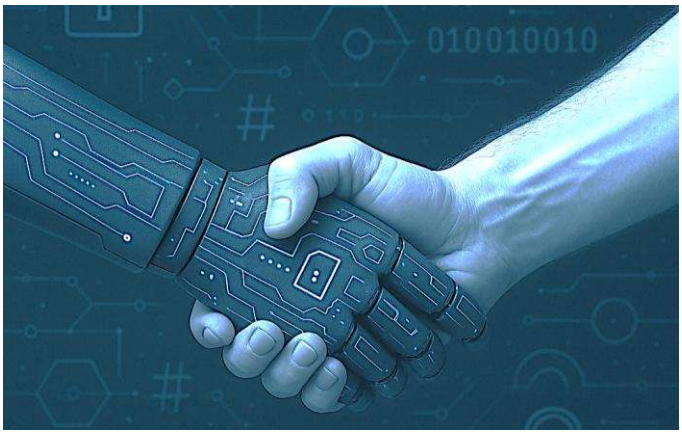
Cryptographic identity binding is your institution's digital passport control, ensuring that every AI agent is who it claims to be, and every action can be traced back to a trusted source. Every agent must not only prove who they are, but also demonstrate they have clear, revocable permission to act on your behalf and only within the boundaries you set.

Whether you're using decentralised wallets, blockchain-based IDs, or next-generation biometric tokens, the goal is the same: to make sure your AI agents don't go off on unauthorised adventures.

As identity-bound AI becomes the new standard, you'll not only keep the fraudsters at bay but also build a foundation of trust that regulators, customers, and partners can rely on.

More recently, I came across an identity platform dedicated to agentic AI: Anonybit[14] provides real-time, decentralised biometric authentication that binds a human's identity to AI agent actions using cryptographically secure tokens. This ensures each agentic action is traceable, auditable, and revocable.

This layered approach proved instrumental in SmartUp's agentic commerce platform[15], where identity-bound AI reduced fraudulent procurement orders by 82% while maintaining sub-second transaction processing. I believe this may be a first in the industry and the promising results will in time be validated through broader industry adoption and independent assessment.



> ***Your AI shouldn't be making decisions for you unless it can prove not just who it is, but that it's genuinely authorised to act for you.***

How Zero Trust Applies to Agentic AI

Zero trust requires that no entity, user, device, or AI agent, is trusted by default. Every access request is continuously verified, with least privilege, time-limited permissions.

- **Multi-Factor Authentication (MFA)** remains essential for agentic AI, especially when approving payments or overriding KYC, using biometric checks or secure hardware tokens to authorise sensitive actions.
- **Role-Based Access Control (RBAC)**, integral to Zero Trust, grants permissions according to defined roles, enforcing least privilege and ensuring AI agents and users access only the data and functions they require.
- **Micro-segmented authorisation** isolates AI capabilities, data stores and execution environments so each agent can access only what it needs, limiting the impact of any breach.
- **Behavioural consistency checks** use (Large Language Model) LLM- powered guardrails to confirm each decision follows ethical guidelines and operational boundaries, preventing manipulation or unintended deviations.

MFA, RBAC, micro-segmentation, and behavioural monitoring should work in concert with each other to create a layered, adaptive zero trust defence fit for the AI age.

Traditional digital identity solutions, such as Okta[18] or Microsoft Entra[19], are widely used for adaptive, risk-based authentication and API security. Whilst there is no evidence of production use for agentic AI, their continuous authentication, conditional access, and micro-segmentation features could be adapted to enforce zero trust for AI agents.

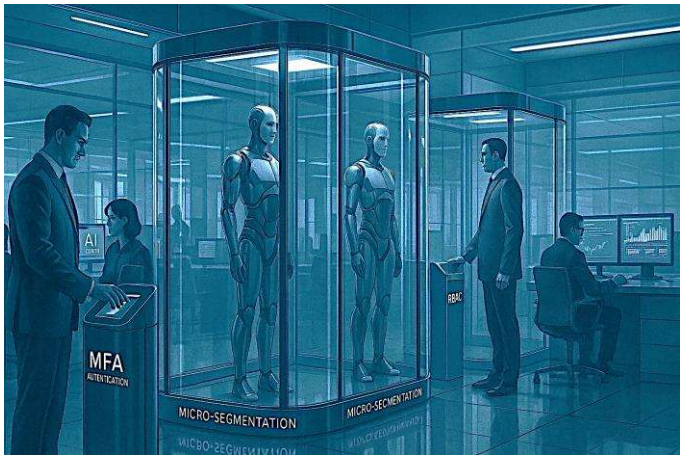
> ***Zero trust isn't a buzzword, it's your best defence against digital mischief. Trust nothing, verify everything, and let your AI earn its privileges, one decision at a time.***

Traditional zero trust segmentation platforms, such as Illumio[20], Guardicore/Akamai[21], and Palo Alto Networks[22], enable organisations to create granular security policies that isolate workloads and data, limiting the impact of breaches and improving overall security posture. They could potentially be adapted to isolate agentic AI processes.

As for **agentic-AI specific commercial solutions**, Anonybit, as previously mentioned, provides biometric-based authentication and authorisation for AI agents, enforcing zero trust and micro-segmentation in production with SmartUp. Another example is Sakura Sky's Vertex AI implementation[23], which restricts payment approval agents from accessing customer KYC databases without explicit need, demonstrating the effectiveness of micro-segmented authorisation in practice.

Zero trust in the age of agentic AI isn't just a security framework, it's your institution's seatbelt, airbag, and satnav all rolled into one. By layering multi-factor authentication, role-based access, micro-segmentation, and behavioural monitoring, financial leaders can ensure their AI agents never wander off course, no matter how fast the landscape changes.

The message is simple: trust nothing by default, verify everything in real time, and always keep your AI on a short, well-audited leash. After all, in a world where algorithms can approve millions in milliseconds, "just enough access, just in time, and just for the right reasons" should be your new mantra.



The Observability Imperative: Continuous Behavioural Monitoring

Continuous monitoring helps ensure AI agents stick to their rules and lets us spot problems quickly. Consistency checks verify that each decision matches the agent's ethical and operational limits, flagging any attempts at hijacking, privilege misuse or unauthorised goal changes.

Intent tracing	Following how an agent's initial instructions lead to its final actions.
Resource consumption profiling	Spotting unusual patterns in API usage.
Context drift detection	Alerting when an agent operates beyond its authorised scope.
Cross-platform audit trails	Linking agent activity across cloud and on-premises systems.

For payment processors, these capabilities are invaluable for swiftly identifying and mitigating risks, such as when a fraud detection AI, compromised by adversarial manipulation or corrupted data, inadvertently approves unauthorised transactions.

Traditional SIEM* platforms (e.g. IBM QRadar[24], Splunk[25], SentinelOne[26], Securonix[27], CrowdStrike[28], Anomali[29]) are not currently deployed for agentic AI but could evolve to support continuous behavioural monitoring of AI agents. The next generation of AI-powered SIEM[30] ("SIEM 4.0" or "agentic SIEM") is expected to integrate agentic analytics, goal consistency validation, and autonomous response, making them suitable for future agentic AI oversight.

** Security Information and Event Management*

As for **agentic-AI specific commercial solutions**, I recently came across these two:

- **HiddenLayer AIDR**[31] provides real-time visibility, guardrails, approval thresholds, anomaly detection, behavioural consistency checks, and audit-ready logs for agentic AI systems, supporting oversight, risk management, and compliance with frameworks like the EU AI Act. It is now available commercially across AWS, Azure, and Google Cloud marketplaces[32].
- **Concentrix**'s agentic observability framework[33] introduces real-time monitoring of AI decision-making processes across the above four dimensions, setting a benchmark for future agentic AI monitoring.

Continuous behavioural monitoring isn't just a technical checkbox, it's the AI equivalent of a dashboard warning light and a vigilant co-pilot, helping you spot trouble before it turns into a headline. By tracing intent, profiling resource use, detecting context drift, and linking audit trails across every platform, you give your AI agents the supervision they desperately need (even if they never ask for a day off).

As agentic AI takes on ever more responsibility, it's not enough to hope your digital workforce is behaving, you need the tools to know, and the confidence to prove it.



> ***In the world of agentic AI,***
> ***trust is good, but continuous, intelligent monitoring is better.***
> ***Don't just hope your AI is on its best behaviour,***
> ***make sure you can prove it, every step of the way.***

Immutable Audit Trails

Immutable audit trails are essential for accountability, regulatory compliance, and forensic investigation. By continuously and cryptographically recording every agent interaction, whether an API call, data access, or transaction, these systems ensure that records remain tamper-proof and trustworthy. Emerging standards now demand audit capabilities exhibiting regulatory-grade accountability, requiring that every AI-driven financial decision be fully transparent and reconstructable, including:



Prompt versioning

Capturing the exact instructions or prompts given to an AI agent at every decision point, ensuring that the original context and intent behind each action can be reconstructed and reviewed.



Tool usage histories

Recording which tools, APIs, or external resources an AI agent accessed or invoked during its operation, providing a clear trail of how each decision was executed and which systems were involved.



Memory state snapshots

Capturing snapshots of an agent's internal state at key moments, so that investigators can understand the data and context influencing each decision.

Such comprehensive auditability allows organisations to trace, verify, and scrutinise every decision an AI agent makes, instilling confidence in regulators, stakeholders, and customers.

I previously mentioned Anonybit, and they provide cryptographically secure, tamper-proof audit logs for every agentic action in production workflows. Google Vertex AI Logging captures prompt versions, tool usage histories, and memory snapshots for AI-driven actions, and could be adapted to support agentic AI auditability in regulated financial environments but is not yet deployed for this purpose.

In a world where AI agents can approve million-pound transactions in milliseconds, an immutable audit trail is your institution's black box and safety net rolled into one. It's not just about ticking a compliance box, it's about knowing, with certainty, who did what, when, and why. Because when the regulators come knocking (and they will), you'll want more than just a good story, you'll want the receipts.

> ***If you can't audit agentic AI, you can't trust it. Make your audit trail as unbreakable as your ambition.***

Adaptive Governance Frameworks

Adaptive governance ensures that security policies and controls for AI agents are continuously refined to address emerging risks and regulatory changes. Automated policy enforcement is central to this approach, leveraging dynamic risk scoring and real-time rollback mechanisms to adjust agent permissions or reverse unauthorised actions, such as restricting trading bots during volatile markets or undoing unintended fund transfers.

Existing enterprise AI governance frameworks, such as Accenture Responsible AI Framework[34] and IBM OpenPages[35], provide dynamic policy engines and risk-based controls and could potentially be adapted for agentic AI. HiddenLayer AIDR, mentioned previously, enables dynamic risk assessment, policy enforcement, and compliance monitoring for agentic AI systems.

Adaptive governance is the difference between letting your AI agents run wild and making sure they know the house rules (and what happens if they break them). With dynamic policy engines, real-time risk scoring, and instant rollbacks, you don't just set the boundaries, you move them as the landscape shifts. Whether it's reigning in a trading bot during market turbulence or undoing an accidental fund transfer, adaptive governance keeps your digital workforce in check, even when the unexpected happens.

> ***Set the boundaries, adjust on the fly, and always be ready to hit undo when your AI gets a little too creative.***

The Governance Gap: *Are Incumbents Addressing the Core Principles?*

A new era of commerce is unfolding as industry leaders, such as Visa, Mastercard, Amazon, and Stripe, deploy agentic AI to transform how we shop and pay.

Mastercard[36] and **Visa**[37] are pioneering tokenisation and agent verification, enabling trusted AI agents to search, select, and securely pay for goods or services on a user's behalf, with robust controls and fraud protection.

Amazon[38] is integrating agentic AI into its ecosystem to deliver seamless, personalised shopping experiences, while Stripe[39] empowers AI agents to automate payments, manage spending, and detect fraud in real time for both consumers and businesses.

Ant International's **Antom** has emerged as a pivotal player, launching what it describes as the world's first comprehensive agentic payment solution in September 2025, enabling AI agents to complete transactions autonomously across both card networks and alternative payment methods whilst partnering with both Mastercard and Visa for tokenised card-based agentic payments[40].

Wells Fargo achieved unprecedented scale by deploying AI agents across all 215,000 employees through Google Cloud's Agentspace platform, automating complex tasks from foreign exchange inquiries to managing over 250,000 vendor contracts[41].

Goldman Sachs crossed another frontier by deploying Devin, an autonomous AI software engineer, marking the first deployment of fully autonomous coding agents on Wall Street that function as "new employees" rather than tools[42].

Across the board, these players are blending advanced security, user control, and interoperability to ensure AI-driven transactions are not just fast and intuitive, but also transparent and trustworthy, heralding a future where digital agents handle everything from party planning to cross-border procurement, all with a single prompt.

But how do they align with the core principles?



cryptographic identity binding



zero-trust architecture



continuous behavioural monitoring



immutable audit trails



adaptive governance frameworks

Cryptographic Identity Binding

Visa and Mastercard are leading with tokenisation: Visa's Intelligent Commerce and Mastercard's Agentic Tokens both ensure AI agents transact only with user-bound, tokenised credentials. Amazon uses biometrics and passkeys for initial authentication of its Nova Act agents, while Stripe demonstrates tokenised credential autofill for agent-driven checkouts. Ant International's Antom has achieved breakthrough integration by partnering with both Mastercard and Visa to enable AI agents to complete transactions autonomously using tokenised card-based credentials whilst maintaining cryptographic safeguards and PCI compliance. Wells Fargo has deployed identity-bound AI agents across all 215,000 employees through Google Cloud's Agentspace platform, demonstrating enterprise-scale cryptographic identity binding in practice. All incumbents are prioritising strong, user-linked digital identities for agentic transactions.

Zero Trust Architecture

Visa and Mastercard both require agent registration, delegated authorisation, and real-time fraud checks, with Mastercard also mandating biometric authentication for agent actions. Amazon and Stripe have not disclosed comprehensive zero trust frameworks for agentic AI, focusing instead on seamless user experience and merchant integration. Among major financial institutions, Wells Fargo's enterprise deployment demonstrates the most advanced implementation, utilising Google Cloud's Agentspace platform to enforce granular access controls and micro-segmentation across 215,000 employees, whilst Goldman Sachs has established role-based access control specifically for its autonomous software engineering agents. However, despite integrating zero trust principles into their enterprise platforms, Bank of America and JPMorgan have not disclosed comprehensive frameworks for agent-to-agent authentication, and whilst Ant International incorporates delegated authorisation into its payment workflows, industry-wide adoption of mature zero trust architectures for agentic AI remains limited.

Continuous Behavioural Monitoring

Mastercard flags unrecognised agentic transactions for user review, and Stripe applies existing fraud detection algorithms to agent-driven payments. Visa and Amazon provide limited detail on real-time monitoring or anomaly detection for agentic AI. JPMorgan has developed comprehensive agentic AI monitoring capabilities, with systems that analyse millions of transactions in real-time achieving 98% accuracy whilst reducing false positives by 95% in anti-money laundering surveillance[43]. Morgan Stanley has achieved 98% adoption of AI systems among advisory teams with enhanced behavioural consistency monitoring that tracks agent decision-making patterns and ensures alignment with defined ethical and operational boundaries[44]. Whilst some incumbents have begun implementing observability frameworks for their agentic AI systems (e.g. Visa's MCP-based monitoring, Mastercard's behavioural biometrics tracking, Amazon's agent workflow tracing), none have yet publicly disclosed full implementation of advanced observability frameworks specifically featuring intent tracing or context drift detection as integrated, production-ready solutions across all their agentic AI operations.

Immutable Audit Trails

None of the major incumbents have publicly committed to providing immutable, cryptographically verifiable audit trails for agentic AI actions. There is a notable absence of detailed, transparent logging of agentic payment decisions across all players, though emerging regulatory requirements under PSD3 mandate that autonomous payment decisions must be cryptographically verifiable and reconstructable with complete decision pathways preserved for ten-year retention periods.

Adaptive Governance Frameworks

Mastercard is collaborating with partners like IBM (with watsonx Orchestrate[45]) to explore governance for B2B agentic workflows, but no incumbent has yet deployed a full adaptive governance framework for agentic AI. Most focus remains on interoperability and ecosystem control rather than dynamic policy enforcement.

Bridging the Gap: From Current State to Best Practice

Agentic AI is rapidly reshaping the financial landscape, with major players deploying advanced technologies to enhance security, streamline payments, and personalise customer experiences. Yet, despite these advances, critical elements such as comprehensive real-time monitoring, tamper-proof audit trails, and dynamic governance remain largely unrealised.

As the sector continues to innovate, the focus is shifting from foundational capabilities to the more sophisticated controls and transparency that will underpin trust and accountability in the next phase of AI-driven finance. While leading financial institutions have made notable progress in cryptographic identity binding in areas such as tokenisation and agent verification, the journey to best practice in agentic AI governance is far from complete. Achieving true resilience and regulatory confidence will require more than incremental improvements, it demands a holistic, forward-looking approach.

To bridge this gap, financial organisations should focus on four key priorities: immutable audit trails, real-time observability, granular zero trust controls, and adaptive governance.

True excellence also means embracing a culture of continuous improvement, regularly reviewing controls, learning from incidents, and collaborating across the industry to shape interoperable standards. As the competitive landscape shifts, those who invest in robust, transparent, and adaptive agentic AI governance will not only satisfy regulators but also earn the trust of customers and partners.

The next competitive frontier in financial services will be defined by institutions that can demonstrate not just technological innovation, but also leadership in governance, transparency, and real-time oversight of agentic AI. Bridging the gap from current practice to best practice is not just a regulatory imperative, it is a strategic necessity for sustainable growth and resilience in the AI-driven future.



Immutable Audit Trails

- Establish tamper-proof, cryptographically signed logs for every agentic decision, including prompt versions, tool usage, and memory states.
- This will not only support regulatory compliance (e.g. PSD3's 10-year retention rule) but also provide the transparency regulators and auditors increasingly expect.

Real-Time Observability

- Implement advanced monitoring capabilities such as intent tracing and context drift detection to identify adversarial manipulation and unauthorised agent behaviour as it happens.
- Leveraging state-of-the-art observability frameworks will be crucial for maintaining trust and operational integrity.

Granular Zero Trust Controls

- Apply role-based access control (RBAC) and micro-segmentation at the API and tool level, ensuring that agentic AI systems operate strictly within their authorised boundaries.
- This minimises risk exposure and limits the potential impact of compromised agents.

Adaptive Governance

- Develop dynamic policy engines that can automatically adjust agent permissions in response to emerging risks, such as market volatility or suspected fraud.
- Integrating real-time compliance checks and automated rollback mechanisms will allow institutions to respond swiftly to evolving threats.

As agentic AI transforms financial services at pace, industry leaders are charting the course, yet their varied strategies expose both notable advancements and enduring shortcomings. For executives seeking to future-proof their organisations and sustain regulatory and competitive edge, a clear grasp of these real-world approaches and their limitations is indispensable.

- **Tokenisation and biometrics are the industry's baseline:**
  - Most institutions now rely on tokenisation and biometrics to secure agentic AI transactions.
  - Visa and Mastercard are at the forefront, using tokenised credentials and biometric authentication to ensure that AI agents act only on behalf of verified users.
  - However, while these measures strengthen security, they do not yet deliver the level of explainability or auditability that regulators and auditors increasingly expect.
- **Strategic Priorities and Ecosystem Approaches:**
  - **Visa and Mastercard are prioritising interoperability and building broad ecosystem partnerships**, working with technology providers to integrate agentic AI into payment networks and ensure seamless, secure transactions across platforms. Their focus is on creating industry standards and supporting agentic commerce at scale, as evidenced by their collaborations with firms like OpenAI, Microsoft, and Stripe[46].
  - In contrast, **Amazon is leveraging agentic AI to deepen platform lock-in**, integrating AI agents into its own ecosystem to deliver highly personalised and automated shopping experiences, but with less emphasis on open standards or cross-industry collaboration.
  - **Stripe is capitalising on its extensive merchant base to drive rapid adoption of agentic AI in payments**, offering tools for automation, spending controls, and fraud detection. However, Stripe's innovation has so far been more focused on user experience and operational efficiency than on advancing governance or transparency.
- **Persistent Industry-Wide Gaps**: despite these advances, significant gaps remain across the sector. Industry leaders have yet to fully implement immutable, cryptographically verifiable audit trails for agentic AI actions. Explainable agentic decision-making and advanced behavioural monitoring are still underdeveloped, leaving organisations exposed to regulatory, operational, and reputational risks.
- **The Next Competitive Frontier**: the institutions that will lead in the coming years are those that invest in robust, transparent governance frameworks and real-time observability for agentic AI, especially in high-stakes domains such as payments and identity verification. The ability to demonstrate not only technological capability, but also proactive oversight and rapid incident response, will become a key differentiator.

While industry leaders such as Visa, Mastercard, Amazon, Stripe, and others, are driving the adoption of agentic AI, their approaches highlight both the progress made and the work still to be done.

For financial executives, the lesson is clear: securing agentic AI is not just about adopting the latest technology, but about building transparent, auditable, and adaptive governance that can keep pace with both innovation and regulation. Those who rise to this challenge will not only satisfy evolving regulatory expectations but will also earn lasting trust from customers and partners, securing a sustainable competitive advantage in the AI-driven future.





The Innovation Dilemma: *The Regulatory Imperative*

The regulatory landscape for agentic AI has shifted dramatically from theoretical frameworks to active enforcement. What were once frameworks struggling to keep pace have crystallised into operational requirements with real deadlines, penalties, and oversight bodies.

The Bank of England's warnings about collective AI behaviour amplifying market shocks are no longer hypothetical concerns, they're driving current supervisory expectations and governance frameworks that financial institutions must demonstrate today, not tomorrow.

The challenge of governing systems that "devise their own strategies, learn from previous experiences, and even collaborate with other AI agents" has moved from regulatory contemplation to active implementation. Traditional compliance approaches haven't just been rethought, they've been rewritten, with new standards now in force across multiple jurisdictions.

Here's what's actually happening:

EU AI Act[47]

The EU AI Act General Purpose AI model obligations became operational on August 2, 2025, with particular implications for autonomous decision-making systems. While the Act doesn't explicitly define "agentic AI," its technology-neutral approach means autonomous systems that can "operate with varying levels of autonomy" fall squarely within scope[48].

Critically for financial services, agentic AI systems are likely to be classified as "high-risk" when deployed in contexts such as credit scoring, fraud detection, or payment processing, precisely the areas where autonomous agents offer the greatest value. The Act's Article 6 classification means these systems must undergo rigorous conformity assessments, implement robust risk management frameworks, and maintain detailed technical documentation of every autonomous decision[49].

The enforcement reality is immediate: General Purpose AI models powering many agentic systems now operate under active regulatory oversight, with the AI Office wielding enforcement powers that apply regardless of where financial institutions are based. For agentic systems built on these foundation models, this creates a compliance chain extending from the underlying AI model through to the autonomous agent's deployment[50].

Most significantly for autonomous systems, Article 14's human oversight requirements present particular challenges. The Act mandates "effective human oversight" for high-risk systems, but agentic AI is specifically designed to reduce human intervention. This creates a regulatory tension that institutions must resolve through what experts dub "redefining oversight": implementing adaptive controls, approval thresholds, and guardrails based on the autonomous system's risk level and behaviour patterns.

The transparency obligations have evolved beyond simple decision logs to encompass autonomous decision pathways, tool usage logic, and the ability to reconstruct how an AI agent arrived at any specific action. Traditional documentation like model cards must be extended to capture the dynamic, goal-driven nature of agentic systems, with the 6% revenue penalty providing a powerful incentive for comprehensive compliance[51].

ISO/IEC 42001:2023[52]

The AI management standard has transitioned from best practice guidance to regulatory necessity. As EU AI Act compliance requirements take effect, the standard's 38 controls for AI-specific threats like data poisoning and adversarial attacks have become the operational baseline for demonstrating regulatory compliance rather than voluntary adoption. The standard's "Plan-Do-Check-Act" methodology is now foundational for financial institutions managing agentic AI in payments and identity verification, particularly as autonomous systems require continuous improvement cycles that can adapt to emerging threats.

PSD3 Proposals[53]

With PSD3 completion expected by year-end 2025, **the explainability requirements for real-time payment systems have evolved from theoretical proposals to concrete compliance obligations**[54]. The cryptographic proof requirements for AI decision-making paths represent a fundamental shift in how autonomous payment systems must document their reasoning. Where the original proposals targeted explainability in real-time systems, the current legislative texts now mandate that every autonomous payment decision must be cryptographically verifiable and reconstructable, with the full decision pathway preserved for the required ten-year retention period.

This directly addresses scenarios like the "£50,000 transfer nobody can explain“: under the emerging PSD3 framework, such transactions would be impossible to approve without maintaining cryptographic evidence of exactly how the AI agent reached its decision, which variables it considered, and how those variables influenced the outcome. The ten-year retention requirement ensures that regulatory inquiries years after deployment can still reconstruct the complete decision logic of any autonomous payment system.

For financial institutions, this represents a compressed timeline where regulatory uncertainty around AI explainability is rapidly giving way to concrete technical requirements that will be directly enforceable across the EU without national transposition delays.

UK's Agentic-Ready Regulatory Evolution[42]

The UK's approach has moved from policy development to operational deployment with specific support for autonomous systems. The AI Live Testing programme launched in summer 2025 and explicitly supports testing of autonomous AI systems in live financial markets. The Supercharged Sandbox provides enhanced computational resources specifically designed to support the intensive processing requirements of multi-agent coordination[55]. In April 2025, the Bank of England published its Financial Stability in Focus: Artificial Intelligence in the Financial System report[56], which identified four critical AI risk areas: core decision-making, market functioning, operational resilience, and cyber threats. The report also established new system-wide monitoring frameworks to address these emerging risks. Furthermore, The Bank’s AI strategy, published August 2025, establishes governance structures that recognise the unique challenges of autonomous systems, including the AI Consortium's focus on systemic risks from correlated AI behaviour[57].

US Federal Regulatory Framework

America’s AI Action Plan, unveiled in July 2025, shifts federal policy toward AI deregulation by directing agencies to remove development barriers and prioritise funding to jurisdictions with lighter AI restrictions. This deregulatory stance accelerates the path for agentic AI pilots and production deployments in finance by lowering compliance hurdles. At the same time, Federal Reserve guidance and research showing that roughly 25% of its workforce roles could benefit from generative AI highlight that US financial regulators are actively considering how autonomous systems will reshape critical banking workflows. Together, these policies create both an incentive to adopt agentic AI at scale and a framework for managing the operational and prudential risks that accompany its deployment[58].

OWASP Agentic Security Initiative[59]

The initiative's identification of 15 critical threats, including memory poisoning and cascading hallucination attacks, has evolved from research findings to operational security frameworks. With financial AI agents now deployed in production environments, these attack vectors represent active threats requiring immediate defensive measures. The memory poisoning scenarios, where AI agents can be manipulated into creating fake employee accounts with admin access, and identity spoofing attacks that manipulate financial AI agents into approving fraudulent transactions, are no longer theoretical, they're documented attack patterns that institutions must actively defend against in their agentic AI deployments.

Cross-Border Regulatory Coordination

Hong Kong financial regulators issued comprehensive AI governance rules in November 2024[60] and August 2025[61], with the HKMA (Hong Kong Monetary Authority) requiring institutions to implement risk-based approaches throughout AI system lifecycles and the SFC (Securities and Futures Commission) establishing four core principles for AI use in investment services[62]. These developments, combined with the Monetary Authority of Singapore continuing its AI governance evolution[63], demonstrate the emerging pattern of coordinated Asian regulatory frameworks that complement rather than conflict with European approaches, creating opportunities for global financial institutions to develop coherent cross-jurisdictional compliance strategies.

The regulatory landscape for agentic AI in finance has moved beyond delicate balancing acts to active enforcement with immediate consequences. What were once foundational tools have become operational requirements with real deadlines, substantial penalties, and cross-jurisdictional oversight. The transformation from theoretical frameworks to concrete compliance obligations means financial institutions can no longer treat AI governance as a future consideration and must make it a present-day operational imperative.

The Bank of England's warnings about AI-driven correlation amplifying financial shocks have materialised into supervisory expectations that institutions must demonstrate today. With the EU AI Act's 6% revenue penalties, PSD3's cryptographic proof requirements, and the UK's live testing programmes all operational or imminent, the regulatory environment has crystallised around transparency, auditability, and human oversight of autonomous systems.

For financial institutions, this represents a fundamental shift from compliance preparation to compliance demonstration. The path forward demands more than adherence to emerging standards; it requires building transparent, auditable AI ecosystems where every autonomous decision can be reconstructed, verified, and defended under regulatory scrutiny. As enforcement mechanisms activate across multiple jurisdictions, success will be defined not by innovation speed alone, but by institutions' ability to prove their AI agents operate within established parameters while delivering competitive advantage.

> ***"The regulatory verdict is in: agentic AI must be autonomous yet accountable, innovative yet auditable, transformative yet transparent, or it won't survive."***

The Latency Trap: *Agentic AI at Scale*

Picture the data centre as a vast railway network where each transaction is a high-speed train requiring precise scheduling, seamless handoffs, and unwavering safety. Traditional hardware, built for earlier workloads, resembles a reliable regional train coping with peak-hour congestion. Agentic AI's demands for persistent memory, instant decision making and extensive parallel coordination simply exceed that design.

Just as today's rail systems deploy purpose-built high-speed lines and advanced signalling, financial institutions must adopt specialised compute platforms to support everything from sub-millisecond trading to autonomous commerce agents that negotiate, order and settle transactions on behalf of customers. Only with infrastructure optimised for agentic scale can organisations deliver the speed, reliability and compliance that the AI era requires.

Why Traditional Hardware Fails

Agentic AI's unique demands, persistent memory for long-term context, ultra-low latency for real-time decisions, and massive parallelism for multi-agent coordination, overwhelm traditional CPUs and GPUs.

For example, a single agentic fraud detection workflow might involve over 10,000 concurrent inferences across structured and unstructured data, while high-frequency trading agents require sub-millisecond latency to avoid catastrophic financial exposure.

How Technology Innovators Are Responding

Technology innovators are in a race to keep pace with the ambitions of agentic AI. From lightning-fast fraud detection to real-time market analysis, leading providers are reimagining hardware and cloud platforms to deliver the speed, scalability, and compliance financial institutions demand

The result? A new generation of intelligent infrastructure, purpose-built to empower secure, auditable, and future-ready AI operations



SPEED **SCALABILITY** **PERFORMANCE**

To achieve **ultra-low latency and real-time processing**, companies like NVIDIA and Graphcore have engineered platforms capable of processing millions of transactions or fraud patterns in parallel. NVIDIA's Blackwell GPUs deliver unrivalled throughput over £11 billion in sales in their first quarter of availability[64] and, when paired with DGX B200 systems[65], power agentic AI workloads such as sub-millisecond fraud detection, market analysis, high-frequency trading, instant payment approvals and automated retail loan underwriting, ensuring rapid auditable decision-making for both customer experience and regulatory compliance. HPE's Private Cloud AI platform extends these capabilities by combining enterprise-grade infrastructure with Blackwell acceleration to meet strict data sovereignty requirements in financial services[66]. Similarly, Graphcore's Bow-2000 IPU[67] enables deterministic scheduling for real-time fraud pattern analysis across millions of retail transactions as well as for continuous liquidity management and portfolio rebalancing in volatile markets, delivering low-latency transparent AI operations essential for modern finance.

Facilitating **seamless data sharing and efficient multi-agent coordination** is another area of innovation. AMD's Instinct MI300X accelerators[68] allow AI agents to share data across complex workflows, reducing delays in scenarios like anti-money laundering (AML) screening of terabyte-scale datasets, or for coordinating customer service bots with fraud detection systems. These accelerators also enable personalised customer service bots to analyse transaction histories and recommend tailored financial products, improving both operational efficiency and financial crime prevention. Cerebras' WSE-3 chip[69] processes entire agentic workflows on a single chip, eliminating inter-device latency. This not only supports compliant audit trails for high-value transactions but is particularly useful for automated loan approval audits with cryptographically signed decision logs.

Cost efficiency, compliance, and secure auditability are also at the forefront of hardware innovation. Intel's Gaudi 3 accelerators[70] prioritise cost efficiency and regulatory compliance for high-volume financial tasks such as payment approvals and fraud detection. Their architecture combines hardware-rooted trust (Intel SGX) with cryptographically secure audit trails, ensuring tamper-proof transaction processing and facilitating regulatory and standards alignment, all while maintaining low operational costs for routine transactions and enterprise-grade security. For cloud deployments, AWS Inferentia[71] reduces cloud inference costs by 50%, enabling scalable deployment of fraud detection bots in retail banking and supporting real-time agentic workflows, such as monitoring transaction patterns for adversarial manipulation, without upfront hardware investments. Google's Ironwood TPU[72] is tailored for large language models and regulatory-grade explainability, making it suitable for training explainable AI models and for EU AI Act-compliant systems. It ensures audit-ready decision-making, such as explainable AI chatbots for customer dispute resolution, balancing energy efficiency with transparency for regulatory scrutiny.

In the meantime, Microsoft's Agent Store, a marketplace for AI agents integrated with Microsoft 365 Copilot, became generally available in May 2025[73]. This initiative allows organisations to access and manage AI agents within existing enterprise environments. For financial institutions, this may facilitate the adoption of **agentic AI for workflow automation and compliance-related tasks,** with potential implications for data governance and auditability depending on implementation and integration practices.

The advancements in technology infrastructure are pivotal in addressing agentic AI's unique demands - speed, scalability, and compliance - while mitigating risks inherent to autonomous decision-making.

By prioritising energy efficiency, cryptographic security, and real-time processing, modern architectures empower financial institutions to deploy agentic AI at scale without compromising latency or regulatory obligations. These innovations ensure that complex workflows, from fraud detection to customer service automation, operate within auditable, tamper-proof frameworks aligned with standards and regulations.

Why Executives Should Know About Hardware Technology

While technology advancements offer new opportunities, financial leaders must also weigh the strategic implications of on-premises versus cloud deployments. On-premises solutions may offer greater control, lower latency, and enhanced data sovereignty, attributes vital for high-frequency trading and sensitive transaction processing. In contrast, cloud-based platforms provide scalability, lower upfront investment, and rapid access to the latest AI capabilities, but may introduce concerns around data residency, vendor lock-in, and regulatory scrutiny. Executives should rigorously assess their institution's risk appetite, compliance obligations, and operational needs when choosing deployment models, ensuring that any move to the cloud does not compromise auditability, security, or long-term flexibility.

Although financial executives rarely interact directly with chip manufacturers like NVIDIA or Intel, the choice of these underlying processors is fundamental to the success of AI deployments. The chips embedded within hardware platforms dictate the speed, efficiency, and scalability of agentic AI applications critical to banking operations such as fraud detection, high-frequency trading, and regulatory compliance. It's not merely about buying hardware, it's about selecting platforms optimised to exploit the unique strengths and limitations of these chips, ensuring that AI workloads run effectively and securely. As one might say, the power of AI lies as much in the silicon beneath as in the software above.

Moreover, **understanding chip-level innovation helps executives anticipate supply chain risks, vendor dependencies, and long-term flexibility, factors that directly influence operational costs and strategic agility.** In a landscape where cloud providers and tech giants increasingly design custom chips to tailor AI performance, staying informed about these developments is essential. Ultimately, infrastructure decisions grounded in hardware insight become a competitive advantage, enabling financial institutions to scale AI responsibly and efficiently.

For financial leaders, the path forward lies in leveraging technology advancements that balance computational power with transparency, enabling institutions to meet evolving regulatory expectations while maintaining customer trust.

As agentic AI reshapes financial services, the fusion of robust infrastructure and adaptive governance will determine whether autonomy becomes a catalyst for innovation or a liability. The future hinges on implementations that not only accelerate AI capabilities but also anchors them to the principles of accountability and human oversight essential for sustainable growth.



> ***"The power of AI lies as much in the silicon beneath as in the software above."***



The Road Ahead: *Embracing Responsible Innovation*

The promise of agentic AI in financial services is nothing short of transformative, but let's not kid ourselves, the road ahead is far from smooth. As institutions race to unlock the benefits of autonomous systems, they're also hitting speed bumps that go well beyond the technical. From legacy systems groaning under new demands to the ever-shifting sands of regulation and the perennial shortage of skilled talent, these challenges are as much about people and culture as they are about code.

For leaders determined to future-proof their organisations, recognising and navigating these hurdles isn't just prudent, it's essential for building trust and staying ahead in the digital era.

TRANSPARENCY

GOVERNANCE

INNOVATION

COLLABORATION

RESILIENCE

Challenges Remain

Despite the remarkable potential of agentic AI, financial institutions face a series of persistent and complex challenges that must be addressed to realise its benefits safely. These obstacles are not merely technical, they touch on culture, risk management, and the ability to adapt to a rapidly shifting regulatory and threat landscape. Understanding these barriers is essential for leaders seeking to future-proof their organisations and maintain trust in the digital era:

- **Legacy integration is a persistent obstacle**, as many banks struggle to reconcile new agentic frameworks with entrenched systems and outdated APIs not designed for autonomous agents.
- **Latency trade-offs are significant**: continuous authentication, behavioural monitoring, and cryptographic checks add processing time, which is especially problematic for latency-sensitive applications like high-frequency trading.
- **Compute overheads is a growing concern**, as agentic AI demands far greater processing power, memory, and storage than traditional AI, driven by multi-agent interactions, persistent memory requirements, and real-time orchestration.
- **Data privacy and security become more complex** as agents require access to vast, sensitive datasets, raising the stakes for robust governance and transparency.
- **Goal misalignment risks** triggering unintended actions that contravene human values, ethical standards, or organisational policies, emerging as agentic AI systems' objectives drift from their programmed intentions due to emergent behaviours in dynamic operational environments.
- **Explainability and trust remain elusive**, as ensuring that agentic AI decisions are interpretable and trusted by stakeholders is a significant, ongoing hurdle.
- **Regulatory compliance is a moving target**, with evolving standards demanding explainability, adaptive controls, and auditable records at scale.
- **Ethical and societal impacts** require regular bias audits and transparency reports, as well as careful consideration of workforce implications and the need to ensure fairness in automated decision-making.
- **Interoperability challenges** include fragmented tokenisation standards, gaps between decentralised identity systems and legacy banking protocols, and diverse legacy infrastructures, complicating seamless integration and scalability.
- **Skills gaps further hinder progress**, with a shortage of professionals trained in both AI governance and complex payment protocols. The IBM May 2025 report "Agentic AI in Financial Services - Opportunities, Risks, and Responsible Implementation"[74] emphasises the importance of AI literacy programmes to address the complex ethical, technical and operational challenges posed by AI adoption and highlights the need for interdisciplinary collaboration and organisational culture shifts to ensure responsible AI implementation.

Yet, for all these hurdles, the rewards of agentic AI remain too significant to ignore. Investment firms are already deploying AI agents to monitor markets, uncover hidden patterns, and optimise portfolios at speeds beyond human capability. In anti-money laundering and compliance, agentic AI is giving teams the tools to respond to financial crime with unprecedented agility.

But real progress will require more than technical innovation: it will demand close collaboration between technologists, compliance experts, and regulators, as well as investment in modernising infrastructure and upskilling teams. Navigating the complexities of interoperability, explainability, and evolving regulation will call for both creativity and discipline.



Building for the Future

The successful adoption of agentic AI in financial services will depend not just on technical prowess, but on commitment to transparency, robust governance, and alignment with human values.

As the pace of innovation accelerates, the ability to anticipate and address emerging risks will define those institutions that lead, rather than lag, in the next era of finance.

Real progress, however, demands more than incremental upgrades or isolated technology deployments.

Financial institutions must take a holistic, forward-looking approach, prioritising cross-functional governance, modernising infrastructure, and investing in the upskilling of teams to bridge the gap between AI, compliance, and payments.

At the same time, policymakers play a critical role in issuing clear guidance, supporting regulatory sandboxes, and fostering industry collaboration to ensure that standards evolve in step with innovation.



> ***Innovation without integrity is just risk in disguise.***

Strategic Imperatives for Financial Institutions

For financial institutions, the following strategic imperatives are critical to unlocking the full value of agentic AI while managing its risks:

- **Establish cross-functional agentic AI governance teams** to oversee deployment, risk management, compliance, and ethical standards.
- **Modernise infrastructure and APIs** to support secure, real-time orchestration and monitoring of agentic workflows, including budget allocation for legacy upgrades and careful consideration of cloud versus on-premises deployments.
- **Strengthen data governance and privacy frameworks** to ensure responsible handling of sensitive financial and personal data accessed by AI agents.
- **Invest in workforce upskilling** in AI governance, explainability, regulatory compliance, and secure software development, including AI-specific engineering, integration, and DevOps practices.
- **Prioritise pilot projects** such as advanced monitoring and audit solutions for behavioural consistency checks (e.g. HiddenLayer AIDR, Concentrix Observability) and cryptographically bound agents (e.g. Anonybit) for high-value transactions.
- **Develop and regularly test incident response plans** tailored to AI-driven processes, ensuring rapid mitigation and transparent communication in the event of failures or breaches.
- **Engage in industry consortia** to help shape and adopt interoperable frameworks for agentic AI security, auditability, and best practice sharing.
- **Communicate proactively with customers and stakeholders** about the role, safeguards, and benefits of agentic AI to build and maintain trust.
- **Define and monitor clear metrics or KPIs** to measure the effectiveness, safety, and compliance of agentic AI initiatives.

By acting on these strategic imperatives, financial institutions can move beyond compliance and truly harness the transformative power of agentic AI. This is not just about deploying new technologies, but about embedding transparency, resilience, and ethical leadership into the very fabric of your organisation.

The institutions that thrive will be those that modernise with purpose, communicate openly, and invest in the skills and safeguards needed for sustainable innovation. After all, in the age of autonomous intelligence, trust isn't just a value-add, it's your greatest asset, and the foundation on which tomorrow's financial leaders will stand.

> ***"Trust isn't just a value-add, it's your greatest asset."***



Policy Pathways for Regulators

Policymakers play a pivotal role in shaping the environment for responsible agentic AI adoption. Key policy considerations include:

- **Issue clear, actionable guidance** on agentic AI accountability, auditability, and explainability requirements. This may include fast-tracking regulatory guidance (e.g. PSD3) on agentic AI audit trails.
- **Support regulatory sandboxes** for controlled piloting of agentic AI solutions and technologies.
- **Facilitate and promote industry collaboration** to develop, adopt, and maintain open, interoperable technical and regulatory standards for agentic AI, including cryptographic identity, data exchange, auditability, and tokenisation, through forums, working groups, and cross-industry initiatives, ensuring compatibility across platforms and reducing fragmentation.
- **Continuously review and update standards** to address emerging risks and technologies.
- **Encourage transparency and public reporting** on agentic AI incidents and governance practices to build trust and accountability.
- **Promote international regulatory alignment** and cooperation to address cross-border AI risks and ensure consistency with global standards.
- **Invest in regulator upskilling and capacity building** to keep pace with technological advances and emerging risks.
- **Encourage ethical and societal impact assessments** for agentic AI deployments to support responsible innovation and public confidence.
- **Strengthen enforcement and supervision mechanisms** to ensure effective compliance and oversight.

Regulators stand at the frontline of shaping a future where agentic AI drives innovation without sacrificing public trust or safety.

By setting clear expectations, fostering industry collaboration, and investing in their own expertise, policymakers can turn potential risks into opportunities for responsible growth.

In this rapidly evolving landscape, the most effective regulators will be those who see governance not as a brake, but as the steering wheel for progress, because smart regulation doesn't slow innovation, it keeps it on course.



> ***"Smart regulation doesn't slow innovation it keeps it on course."***

Take the Wheel

So, as AI agents race ahead, navigating markets, flagging fraud, and making decisions in the blink of an eye, the real question isn't whether we're on the journey, but whether we're in control of the itinerary. The future of finance won't be shaped by those who simply automate, but by those who demand transparency, accountability, and a healthy dose of human oversight at every turn.

Yes, there will be wrong turns, unexpected detours, and the occasional flat tyre as we integrate agentic AI into the heart of our institutions. But with robust governance, cross-functional teamwork, and a willingness to modernise both mindsets and infrastructure, we can keep our hands firmly on the wheel. For financial leaders, this means not just keeping pace with innovation but setting the pace, ensuring that AI autonomy always serves the interests of trust, compliance, and sustainable growth.

Let's not settle for being passengers in the age of agentic AI. Map in hand, eyes on the road, let's drive towards a future where automation is not just fast, but fair and fully accountable. The journey is just beginning, let's make sure we're in the driver's seat.

The future isn't just automated, it's transparent and accountable. Let's steer it that way.



> ***"The real question isn't whether we're on the journey, but whether we're in control of the itinerary."***



REFERENCES

1. https://www.linkedin.com/pulse/just-out-uk-finance-annual-fraud-report-2025-what-latest-neira-jones-8ozre/
2. https://www.ibm.com/think/insights/scale-responsible-agentic-ai
3. https://kpmg.com/uk/en/insights/ai/cdd-and-generative-ai-risks-of-kyc-fraud.html
4. https://thehackernews.com/2025/08/anthropic-disrupts-ai-powered.html
5. https://www.bankofengland.co.uk/financial-stability-in-focus/2025/april-2025
6. https://www.bbc.co.uk/travel/article/20240222-air-canada-chatbot-misinformation-what-travellers-should-know
7. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/how-agentic-ai-can-change-the-way-banks-fight-financial-crime
8. https://www.forbes.com/councils/forbestechcouncil/2025/05/14/understanding-and-controlling-agentic-ai-security-risks/
9. https://www.yoti.com/
10. https://www.lfdecentralizedtrust.org/projects/hyperledger-indy
11. https://trinsic.id/
12. https://www.evernym.com/
13. https://www.civic.com/
14. https://veramo.io/
15. https://www.proofspace.id/
16. https://www.anonybit.io/press-releases/secure-agentic-workflows/
17. https://www.biometricupdate.com/202505/agentic-ai-needs-to-be-bound-to-human-identity-anonybit-offers-a-solution
18. https://www.okta.com/
19. https://www.microsoft.com/en-gb/security/business/microsoft-entra
20. https://www.illumio.com/
21. https://www.akamai.com/products/akamai-guardicore-segmentation
22. https://www.paloaltonetworks.co.uk/zero-trust
23. https://www.sakurasky.com/blog/ai-agents-rbac-vertexai/
24. https://www.ibm.com/qradar
25. https://www.splunk.com/
26. https://www.sentinelone.com/
27. https://www.securonix.com/
28. https://www.crowdstrike.com/
29. https://www.anomali.com/capabilities/siem
30. https://www.detectionatscale.com/p/gen-4-siem
31. https://hiddenlayer.com/innovation-hub/how-to-secure-agentic-ai
32. https://hiddenlayer.com/innovation-hub/hiddenlayer-expands-security-for-ai-solutions-to-major-marketplaces
33. https://www.concentrix.com/services-solutions/agentic-ai/agentic-ai-observability/
34. https://www.accenture.com/gb-en/services/data-ai/responsible-ai
35. https://www.ibm.com/products/openpages
36. https://www.linkedin.com/pulse/mastercard-agent-pay-ai-gets-license-spend-mastercard-2ajcc/
37. https://www.businesswire.com/news/home/20250430580204/en/Find-and-Buy-with-AI-Visa-Unveils-New-Era-of-Commerce
38. https://www.pymnts.com/amazon/2025/customer-first-agentic-ai-next-amazon-lays-out-its-2025-road-map/
39. https://www.smartcompany.com.au/technology/stripe-ai-payment-agents-e-commerce-demonstration/
40. https://www.leaprate.com/financial-services/fintech/ant-internationals-antom-launches-agentic-payment-solution/
41. https://cloud.google.com/blog/topics/financial-services/wells-fargo-agentic-ai-agentspace-empowering-workers
42. https://www.cnbc.com/2025/07/11/goldman-sachs-autonomous-coder-pilot-marks-major-ai-milestone.html
43. https://www.klover.ai/jpmorgan-ai-strategy-chasing-ai-dominance/
44. https://www.aicerts.ai/news/morgan-stanley-ai-adoption-driving-real-returns/
45. https://www.digitalcommerce360.com/2025/05/06/visa-mastercard-ai-agentic-commerce/
46. https://www.pymnts.com/news/artificial-intelligence/2025/visa-mastercard-paypal-fuel-agentic-ai-commerce-boom/
47. https://artificialintelligenceact.eu/
48. https://kennedyslaw.com/en/thought-leadership/article/2025/agentic-ai-what-businesses-need-to-know-to-comply-in-the-uk-and-eu/
49. https://www.mhc.ie/latest/insights/rise-of-the-helpful-machines
50. https://www.loyensloeff.com/insights/news--events/news/the-second-wave-of-ai-act-provisions-enters-into-force/
51. https://hiddenlayer.com/innovation-hub/governing-agentic-ai/
52. https://www.iso.org/standard/42001
53. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52023PC0366
54. https://vinciworks.com/blog/psd3-gets-real-what-the-eus-new-payment-rules-mean-for-you/
55. https://www.fca.org.uk/firms/innovation/ai-lab#section-ai-live-testing
56. https://www.bankofengland.co.uk/financial-stability-in-focus/2025/april-2025
57. https://www.bankofengland.co.uk/about/governance-and-funding/staff-codes-and-policies/the-banks-ai-strategy
58. https://www.faegredrinker.com/en/insights/publications/2025/8/americas-ai-action-plan-and-more-developments-for-the-tech-financial-services-insurance-and-health-care-sectors
59. https://genai.owasp.org/2024/12/15/announcing-the-owasp-llm-and-gen-ai-security-project-initiative-for-securing-agentic-applications/
60. https://brdr.hkma.gov.hk/eng/doc-ldg/docId/20241107-1-EN
61. https://brdr.hkma.gov.hk/gb_chi/doc-ldg/docId/getPdf/20250825-1-TC/20250825-1-TC.pdf
62. https://apps.sfc.hk/edistributionWeb/gateway/EN/circular/intermediaries/supervision/doc?refNo=24EC55
63. https://www.mas.gov.sg/publications/monographs-or-information-paper/2024/artificial-intelligence-model-risk-management
64. https://www.aol.com/finance/nvidia-sells-11-billion-next-215339924.html
65. https://www.NVIDIA.com/en-gb/data-center/dgx-b200/
66. https://datacentre.solutions/news/70204/hpe-unveils-new-ai-factory-solutions-built-with-nvidia
67. https://docs.graphcore.ai/projects/bow-2000-datasheet/en/latest/index.html
68. https://www.amd.com/en/products/accelerators/instinct/mi300/mi300x.html
69. https://www.cerebras.ai/chip
70. https://www.intel.com/content/www/us/en/products/details/processors/ai-accelerators/gaudi.html
71. https://aws.amazon.com/ai/machine-learning/inferentia/
72. https://blog.google/products/google-cloud/ironwood-tpu-age-of-inference/
73. https://devblogs.microsoft.com/microsoft365dev/introducing-the-agent-store-build-publish-and-discover-agents-in-microsoft-365-copilot/
74. https://www.ibm.com/downloads/documents/gb-en/12f5a71117cdc329

GET IN TOUCH

I hope you've enjoyed this report and found it of value, as producing it certainly clarified things in my own mind. If you have time, I'd love to hear your feedback!

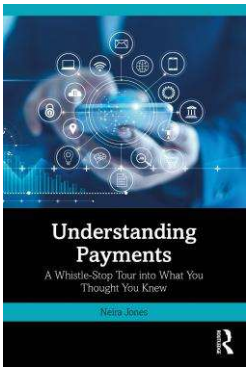
Interested in learning more? I'd love to help you decode the payments world! My keynotes, training and books make payments and fintech fun and accessible. Whether you need consulting, expert advice or training, I'll support your journey with 25+ years of experience in financial services. Let me transform your payment knowledge with a personal touch and a bit of humour!

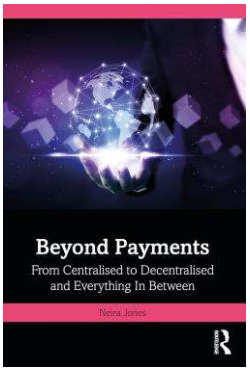
My e-learning platform: https://neirajones.thinkific.com/

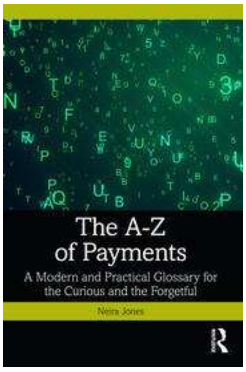
Book me for your event

I regularly address global audiences, in person or virtually, with thought-provoking insights on payments, fintech, and related topics. From intimate boardrooms to bustling conference halls, I transform complex industry concepts into lively, accessible discussions.
Need a keynote that educates whilst entertaining? I deliver presentations that resonate long after the applause fades!

My books







Need help?

With over 25 years in financial services and technology, I'm passionate about change through innovation and partnerships, constantly demystifying the hype around industry trends. I advise organisations of all sizes on payments, fintech, regtech, cybersecurity, and regulatory matters. As a strategic board advisor and non-executive director, I take immense pleasure in collaborating with innovative companies, catalysing their growth through pragmatic guidance and industry savvy.

