TRUST IN THE AGENTIC ECONOMY

GOVERNING AI AGENTS OPERATING UNDER DELEGATED AUTHORITY: IDENTITY, MANDATES AND CONTROLS



KINEXYS
BY J.P.Morgan

AI agents act on behalf of clients with materially less human involvement at each step, and there is a need for a new trust and security infrastructure to support a safer and more secure performance of such transactions.

KEY TAKEAWAYS:

1. Agents change what is transacted, gives rise to a new category of high-frequency, micro-value "enablement payments" that agents make in order to do their work.
2. More fundamentally, agents change how any action must be governed. Delegated authority is not new, but consequence should still fall on the principal, not the AI agent. Trust and security infrastructure is therefore paramount to protect the principal by keeping agent actions within the authority granted, and giving institutions the means to better govern agents and safeguard for unintended actions.
3. Trust and security in agents is built in layers: who the agent is, what it is authorized to do, whether each action is checked before value moves and a record of what happened.
4. A proof of concept with Hyundai Card shows the framework is buildable today: an AI treasury agent, with a verified identity and a principal-approved mandate, executed FX funding transactions through bank-enforced controls and full traceability. The binding constraint is agreement on common standards, not technical feasibility.

ACKNOWLEDGEMENTS

AUTHORS

WEE KEE, TOH
Global Head of Insights,
Kinexys by J.P. Morgan

MALCOLM NG
Product Manager,
Kinexys by J.P. Morgan

FAYE ALSAADOON
AI & Shared Services Lead,
Kinexys by J.P. Morgan

FINLEY MCDIARMID
Product Manager,
Kinexys by J.P. Morgan

CONTRIBUTORS

UMAR FAROOQ
Co-Head of Global Payments,
J.P. Morgan

OLIVER HARRIS
Global Head of Kinexys,
Kinexys by J.P. Morgan

SAI VALIVETI
Head of Product Engineering,
Kinexys by J.P. Morgan

DEBIDUTTA SAMANTARAY
Principal Enterprise Architect,
Kinexys by J.P. Morgan

ALOYSIUS LIM
Lead Software Engineer,
Kinexys by J.P. Morgan

SWAGNIKA BHATTACHARJEE
Software Engineer,
Kinexys by J.P. Morgan

HEEJUNG NAM
Division Head, Payment Business Division,
Hyundai Card

SUNG MIN, CHUNG
Department Head, Payment Business,
Hyundai Card

INTRODUCTION

AI agents are increasingly being designed to act autonomously on behalf of clients. They negotiate, decide, and transact with far less direct human involvement at each step. This shift is particularly apparent in the context of payments, because payments convert autonomous decisions into real-world commitments and transfers of value. However, the underlying change is bigger than payments: it is about how autonomous software acting under delegated authority is identified, authorized, and controlled.

This shift plays out in two distinct ways. First, agents are changing what gets transacted - giving rise to novel categories of activity, such as the high-frequency, machine-to-machine "enablement payments" agents make in order to do their work. This includes data, compute, search, and payments to other agents. Second, and more fundamentally, agents are changing how any transaction, new or familiar, must be governed and ultimately validated: who is the agent, what authority does it carry, and what stops it from acting outside that authority. We believe that identity, mandates, and controls address this second shift.

Throughout this paper we refer to the "triple", comprising, principal, agent, and mandate, as the unit of accountability for any agentic action. Every governed action must specify who granted the authority (the principal), which specific agent identity exercised it (the agent), and the scope, limits, and expiry bounded that constrained it (the mandate). Identity establishes which agent acted; delegation and authorization define the mandate, and the principal who granted it; and controls ensure every action remains consistent with the full triple.

This paper introduces a conceptual framework for governing agentic AI built on five elements: verifiable identity and explicit mandates, deterministic and AI-based controls, fit-for-purpose settlement, comprehensive attribution and reporting, and machine-native trust. The framework applies to any domain in which agents act under delegated authority, from treasury and procurement to fleet management and research.

This paper is organized in three parts. The first sets out the ecosystem-wide position: why agentic activity changes how transactions are governed, and the framework needed to govern them. The second then tests that framework in practice through a proof of concept (POC) developed by Kinexys by J.P. Morgan with Hyundai Card, in which a client-instructed AI agent performs treasury operations within a bank-enforced identity, mandate, and control layer. The third draws on both the framework and the POC to outline the path forward and the next steps.

TWO SHIFTS: WHAT'S BEING TRANSACTED, AND HOW IT'S GOVERNED

SHIFT ONE: A NEW CATEGORY OF TRANSACTION

Principal payments are the core actions an AI agent carries out on behalf of its principal, such as the purchase, settlement, or trade that falls within the client-approved mandate. In substance, these transactions may look much like today's transactions: the nature, value, and counterparties are the same, and only the actor is new. Traditional delegation is supported, in part, by the fact that human delegates can be identified, supervised, and held responsible within established legal, contractual, regulatory, and governance frameworks. Absent new frameworks, AI agents, by contrast, do not themselves bear legal or practical consequences. Any accountability must therefore attach to the relevant human or legal person and be embedded into infrastructure. The central governance question is therefore whether the (principal, agent, mandate) triple can be verified for each action: did the agent hold a mandate from this principal, and is it acting within that mandate's scope?

Enablement payments are the payments an agent makes to be able to perform the principal transaction, such as for search, data, API access, compute, and payments to other agents. These are the genuinely new category: high-frequency, micro-value payments fanned out across many machine and agent counterparties, creating new challenges around unit economics, attribution, and control.

SHIFT TWO: A NEW BASIS FOR GOVERNANCE

Regardless of whether a payment is a principal payment or an enablement payment, whether familiar or new, it is now initiated by an autonomous AI agent rather than a human. This changes what it takes to govern it safely.

Delegated authority is not new. Firms have long authorized employees and other representatives to act on their behalf within defined limits, without requiring prior approval for each individual action. What is different with AI agents is that the delegate does not have its own self-interested incentive to avoid the consequences of a poor decision. A human delegate may be subject to professional, financial, reputational, contractual, employment, or regulatory consequences, and that potential exposure can shape behavior even when no specific rule or approval step applies. In that sense, human accountability operates as an embedded control: it is continuous, context-sensitive, and capable of influencing judgment in situations not expressly addressed by policy. It is the reason firms can delegate broadly to people and rely on judgment to fill the gaps.

An agent has no such stake. It does not experience the consequences of its decisions; consequence attaches entirely to the principal, and in some configurations to counterparties who had no relationship with the principal at all. Governance must therefore be extended and encoded in the identity, mandate, and enforced controls, because none of it can be assumed to reside in the agent. The combination of extended governance, clear enterprise-wide allocation of legal, contractual and operational responsibilities and processes for auditability of every AI agent action, supports a more trusted and scalable agent-to-agent experience.

The following dimensions capture this shift:

- Scale, speed and granularity: high-frequency, micro-value activity across many counterparties
- Machine-speed decisioning: no human deliberation or "this feels wrong" pause before execution
- Non-human counterparties: machine-to-machine interaction with APIs, services, and other agents
- Programmatic communication: identity, intent, and terms exchanged through structured protocols, not human-readable contracts

- Delegated identity: agents act on a borrowed, scoped mandate rather than a fixed personal identity
- Indirect link to instruction: actions can sit several autonomous reasoning steps removed from any explicit human instruction

These dimensions are why identity, mandates, and controls, and not faster rails, are the central governance challenge of agentic AI.

THE FRAMEWORK: IDENTITY, MANDATES, AND CONTROLS

Addressing these risks is a natural extension of the role financial institutions already play. Accountability that once rested with the human delegate now also applies to the actions of the AI agents acting on their behalf. This requires existing legal frameworks to be preserved and infrastructure safeguards to be built, so that responsibility stays with the right person or legal entity. The trust, identity, controls, and settlement mechanisms that enable clients to transact safely today are precisely what agentic activity requires, adapted and reimagined for a delegated machine actor rather than a human one. The examples below are drawn from payments and finance, but the concepts apply to any agentic use case.

VERIFIABLE AGENT IDENTITY AND DELEGATED MANDATES

Addressing the authorization risk means making the (principal, agent, mandate) triple provable at the point of action, for both principal and enablement flows.

The model contemplated here achieves this through each agent being provisioned with a verifiable cryptographic identity, and each agent identity being anchored to the institution's verified business identity (KYB), such that the counterparty can verify who the agent is and who it acts for.

This enables a shared trust-anchor model that mutually recognizes issuers, registries, or credential schemes, such that verification performed by one participant is meaningful to another. Each mandate, including details like permitted transaction types, amounts, counterparties, time windows, escalation conditions, is encoded with explicit scope, limits, and expiry, binding every action to the (principal, agent, mandate) triple. An immutable, permissioned audit log records every identity assertion, mandate change, and agent action, giving principals the tooling to grant, scope, review, and revoke delegated authority.

DETERMINISTIC AND AI-BASED CONTROLS AND GUARDRAILS

Agentic activity delivers most of its value where in-scope actions are processed straight-through, with exceptions subject to escalation or human review. Controls must therefore be encoded in the system and enforced before execution. This is not because human judgment has no place, but because it cannot be applied to every action at machine speed and volume. The objective is to minimize routine human intervention while deliberately preserving it for the exceptions that warrant it. Given this, deterministic as well as AI-based controls and guardrails are set up throughout the end-to-end agent to agent interaction, including:

- Deterministic controls, enforced by policy rather than judgment
- Pre-authorization: encoded, policy-based authorization, spending limits, and automated approval rules applied before execution
- Runtime containment: real-time budgets, rate limits, circuit-breakers, and kill-switches to contain machine-speed error propagation and runaway loops
- Escalation thresholds that route out-of-policy (mandate) actions back to a human or the principal for approval, defined across both pre-authorization and runtime containments
- AI-based review of each action against the agent's original mandate to generate a confidence score on whether the action is permitted, with the assessment validated against trusted independent data sources rather than relying solely on information provided by the agent

FIT-FOR-PURPOSE SETTLEMENT

Given the cost of processing traditional payments transactions, the high-frequency, micro-value profile of enablement payments breaks traditional rail economics. This calls for settlement that can settle low-value transfers economically, execute atomically and conditionally without manual intervention, and be consumed directly by machines.

Enablement transactions in any agentic domain, such as data access, compute, agent-to-agent services, demand settlement that can execute atomically and at machine speed, without manual intervention. The framework must accommodate fit-for-purpose settlement that matches the frequency and granularity of agentic activity.

ATTRIBUTION, RECONCILIATION, AND REPORTING

Tracing a long tail of machine actions back to the task that caused them, and giving principals visibility over it, is essential for operational management and governance, in payments and beyond. This creates a need for infrastructure that can make agent activity observable, attributable, and auditable at scale, including:

- WHY - Why was there an agent action: end-to-end logging and traceability tying every agent action back to the originating principal task (cost-to-task attribution)
- WHAT - What did the agent do: real-time monitoring and dashboards purpose-built for high-volume, high-fan-out machine activity
- HOW - How did the agent go about it: reconciliation and reporting tooling giving principals clear visibility into what their agents are doing, with whom, and why

TECHNICAL SECURITY AND MACHINE-NATIVE TRUST FOR AGENT-TO-AGENT INTERACTION

The preceding sections address institutional trust, which concerns whether an agent's authority can be established, and its activity accounted for. This section addresses technical trust, which is the cryptographic and protocol-level assurance that holds against an active adversary, not merely against accidental error. Both are necessary, and they fail in different ways.

Non-human counterparties and input-driven decisioning create a new attack surface that needs machine-native trust and defenses, extending beyond point-in-time verification to continuous, network-wide assessment of agent behavior. Addressing these requires trust and security mechanisms designed for autonomous, machine-to-machine interactions, including:

- Injection-resistant authorization and verified data-source requirements to defend against prompt injection, poisoned data, and adversarial inputs
- Hardened, authenticated agent-to-agent protocols with programmatic verification of counterparty identity, intent, and terms
- Machine-native trust and onboarding for non-human counterparties using verifiable credentials, registries, and reputation signals
- A network-wide trust and reputation layer that observes agent behavior, flags anomalies and mandate violations, and propagates signals across participants

CASE STUDY: AGENTIC TREASURY WITH HYUNDAI CARD

The preceding sections describe the framework in the abstract. This section shows how it operates in a working transaction flow, using a proof of concept built with Hyundai Card. The use case is treasury and payments, but the controls demonstrated can be extended to any domain in which an agent acts under delegated authority.

WHY THIS CASE

In collaboration with Hyundai Card, Kinexys by J.P. Morgan has implemented a proof of concept in which a client-instructed AI agent performs autonomous treasury operations (checking balances, sourcing competitive FX quotes, and executing funding transactions) against a bank-enforced identity, mandate, and control layer. The POC is deliberately narrow in scope and runs in a test environment that is integrated with back-end systems; it is not a live product, and no real value is transacted. Its purpose is to illustrate how the (principal, agent, mandate) triple can be made provable at the point of action in a real transaction flow, and how key questions a bank may seek to address when an agent moves money, namely identity, authority, control, and traceability, can be answered deterministically and at machine speed.

The use case is a currency conversion to fund an operating account, which is considered a principal transaction, as described in section "Shift One: A New Category of Transaction". The nature, value and counterparties are entirely familiar, and only the actor is new. It is therefore a direct test of Shift Two: the governance burden comes not from the type of transaction, but from the autonomy of the agent actor initiating it.

THE SCENARIO

Hyundai Card deploys a treasury management agent whose standing objective is to maintain a minimum balance in its Frankfurt EUR account, funding any shortfall by selling GBP from a named London account. The agent checks the balance on a defined cycle. When the balance falls below the target, it requests quotes from a panel of three banks, selects the most competitive, and instructs the winning counterparty to execute. Quote selection is an internal client step, so only the winning counterparty is asked to execute.

The Hyundai agent plans, compares options and submits requests. It initiates the payment request, and execution sits with a separate agent operated by Kinexys. This separation between the agent's decision-making and the authority to execute provides an infrastructure-level control, ensuring that the agent's reasoning informs the action without being the sole basis or authority for execution. This is critical under the adversarial conditions described in section "Technical Security and Machine-Native Trust for Agent-To-Agent Interaction".

IDENTITY AND MANDATE IN PRACTICE

The principal generates a cryptographic keypair for the agent and registers the public key with Kinexys, alongside an agent identifier and the legal entity it is acting for. The agent authenticates itself and signs outgoing messages with its private key. Identity is therefore not asserted by the agent in conversation; it is verified cryptographically against a registration anchored to the client's verified business identity.

Authority is carried in a separate, machine-executable mandate. The agent drafts a proposed version of the mandate against a published schema, which the principal reviews and approves on the Kinexys Portal. Kinexys binds the approved mandate to that specific agent public key and evaluates it deterministically at runtime. The mandate sets out the delegated authority of the agent, providing a defined list of permitted actions, escalation conditions, and guardrails applicable to the agent's activity within the POC.

Two properties of this design are relevant to the wider industry discussion. First, the mandate is drafted by the agent but has no force until the principal approves it, so the agent can propose the shape of its own authority but never grant it. Second, the mandate carries both a machine-evaluated rule set and a plain-English statement

of intent that the principal reads before approving.

Independently of the mandate's contents, the bank always verifies the agent's identity, that the mandate is approved and within validity, and that every account named belongs to the principal.

CONTROLS AT RUNTIME: FIVE CHECKS, THREE OUTCOMES

When a request reaches the Kinexys agent, it passes through a defined series of checks before anything is released for execution:

On every request, the bank runs the following checks before any money moves:

- Is the agent genuine? Verification that the agent credential is valid and current
- Who is it acting for? Confirmation of which principal the agent is bound to
- Is it within the mandate? Deterministic mandate evaluation against key parameters including the currency pair, transaction size, limits, and operating window
- Does it make sense? An AI-based confidence assessment of whether the request is coherent given the agent's stated objective and whether anything about it is unusual
- Is it on the record? The goal, plan, action, check results, and outcome are recorded to an immutable audit trail

Only when all checks pass is the instruction released for execution.

Every request resolves to one of three outcomes. A request within all parameters and below a principal-defined escalation threshold executes automatically. A request within all parameters but above the threshold is escalated to the principal for approval. A request outside any approved parameter is denied, whether it relates to a different account, an unapproved currency pair or transaction direction, or an inactive authorization. The POC exercises all three. In the primary flow, the agent finds the EUR balance below target, requests quotes, and executes with the winning bank, with the transaction clearing all checks listed above. In the deny flow, the agent is instructed to transact in a currency that was never included in the approved pairs; Kinexys blocks it at mandate verification, and the agent is told the action is not permitted. The remediation path would be an updated mandate from the principal, not a retry. In the escalate flow, a funding amount above the autonomous threshold is held, surfaced in the portal with the triggering rule and the action on hold shown explicitly, and released only once the principal approves.

This delivers the outcome described in section "Deterministic and AI-Based Controls and Guardrails": routine activity is straight-through, and human judgment is preserved for the exceptions that warrant it rather than spent on every action.

ATTRIBUTION AND TRACEABILITY

The POC includes an operational trace view that reconstructs a transaction end-to-end : a unique trace identifier, current status, initiating and processing agents, the principal and legal entity, the mandate identifier and version, timestamps, currency pair and amount, accounts involved, and a chronological journey with each control check marked. Mandate versions are retained so that a historical transaction can be evaluated against the authority that was actually in force at the time, and mandate states are kept explicit: draft, pending signature, active, expired, superseded, and revoked.

The design intent is that if a regulator, auditor, or treasurer asks why something happened, the answer already exists rather than needing to be reconstructed after the fact. Whether this intent is realized in a given live deployment will depend on operational implementation and the retention and access policies of the relevant participants. The constraints on the agent matter just as much: states such as requested, quoted, escalated, approved, executed, and settled are kept strictly distinct, and no action may be reported as complete without runtime confirmation from the executing counterparty. Without that rule, an agent could describe an intention as an outcome, and every downstream record would inherit the error.

WHAT THE POC DEMONSTRATES, AND WHAT IT DOES NOT

The POC illustrates, within the parameters of a controlled test environment, that the framework described is capable of being implemented using currently available technology: a cryptographically identified client agent, a principal-approved machine-executable mandate, deterministic and AI-based pre-execution controls, and end-to-end attribution can be assembled into a working agent-to-agent transaction flow integrated with back-end systems for a defined use case.

The boundaries should be equally clear. The scope is deliberately confined to a single funding objective, a single currency pair and direction, and two named accounts, prioritizing end-to-end testing on a simple use-case. The agent is deployed within the bank's own agent-as-a-service environment, which allows the agent-to-agent interaction to be tested without first solving cross-institutional network connectivity and governance.

The POC tests only a single agent-to-agent interaction based on a single bank-client relationship, and not a multi-agent multi-bank multi-client environment. A mandate approved with one institution is not currently meaningful to another, and the shared trust-anchor model described remains an open industry question rather than a solved one. That gap is the substance of the standardization discussion set out in the Path Forward and Next Steps sections.

GOVERNANCE, EXCEPTIONS AND ACCOUNTABILITY

Agentic systems will inevitably encounter exceptions and failure scenarios. The POC explores some of these scenarios and determines how they can be addressed through technical and operational controls. As agentic systems mature, an increasing number of known failure conditions can be managed through predefined controls rather than ad-hoc intervention, helping to improve consistency, resilience, and trust.

The POC implements a range of mechanisms intended to manage these different scenarios, including agent authentication, principal-approved mandates, approval workflows and auditable decision trails. Transactions that fall outside an approved mandate, exceed predefined limits, or fail specific control checks can be denied or escalated rather than executed autonomously. By testing both expected and unexpected outcomes, the POC helps identify where controls can be strengthened.

However, it is unlikely that every possible scenario can be anticipated in advance. Governance therefore remains an important additional layer, providing oversight and accountability for situations that fall outside established controls. As additional scenarios are explored, the range of conditions that can be managed through technical and operational controls can continue to expand, reducing reliance on governance interventions over time.

While the POC does not establish a definitive legal or liability framework, it provides a practical environment in which to explore how controls, governance and accountability mechanisms can work together to support trust in the agentic economy.

PATH FORWARD: THE AGENTIC INFLECTION POINT

The case study suggests that the framework is capable of being implemented using existing technology and indicates that a key remaining constraint is agreement on common standards rather than technical feasibility. That reframes the question for the industry.

The shift to agentic AI is accelerating rapidly. AI platforms, card networks and fintechs are racing to enable agents to transact autonomously. The central question for the industry, and for banks, is not whether agents will transact, but how much institutions should remain in the transaction flow and maintain governance without upsetting well established liability frameworks.

The framework outlined above on identity, mandates, controls, and trust, borrows concepts from traditional banking and payments but is not a payments-specific construct. It is intended to describe foundational infrastructure considerations that an agentic economy is likely to require, whether the transaction is a treasury settlement, a connected vehicle payment, a research agent accessing premium data, or machine-to-machine commerce between independent agents. Establishing this foundation on an industry-wide basis is how financial institutions can maintain their role as trusted intermediaries in an economy where principals, agents, and autonomous counterparties interact at machine speed.

The urgency is strategic. Agents are likely to gravitate toward rails, interfaces, and trust models that are designed for autonomous interaction, where governance, controls, and verification are embedded into the transaction process. Infrastructure that is not machine-native may be designed out of the flow entirely. By establishing identity, mandate, and control frameworks today, for payments, and by extension for other agentic transaction types, banks may be better positioned to serve clients in an economy where machines act with delegated authority on behalf of their principals.

This is a shared industry challenge. Standardization around verifiable identity, portable mandates, and interoperable controls will benefit all participants. The path forward requires convening clients, regulators, ecosystem partners, and technology providers to establish the shared infrastructure that agentic AI demands.

NEXT STEPS

A fundamental shift is underway in the economy: autonomous agents are increasingly acting with delegated authority across industries. We plan to expand on the POC with Hyundai to explore more complex multi-agent multi-party scenarios to understand the application of agentic payments across a variety of use-cases spanning connected vehicles and fleet management.

The institutions that develop trusted frameworks for identity, mandates, controls, attribution, and machine-native trust may be well positioned to serve as intermediaries in this emerging economy. The opportunity is immediate, and the window to establish common standards is now.

The next step is industry collaboration: establishing shared standards around verifiable agent identity, portable mandates, and interoperable controls. The path forward is not a single bank's or fintech's responsibility, but a shared infrastructure challenge. By working together to establish these foundations, the industry can build an agentic economy that is safe, efficient, and trustworthy.

The POC points to where that collaboration is most needed. A mandate approved with one institution is not yet meaningful to another, so a shared trust-anchor model is required. The single bank-client interaction tested must extend to multi-agent, multi-bank, multi-client environments, which in turn depends on cross-institutional network connectivity and security governance. We invite clients, ecosystem partners, and regulators to engage with us on these foundations.

DISCLAIMER

Some of the solutions described in this paper are not live product offerings and there is no guarantee that J.P. Morgan will offer these solutions. The information set forth herein has been obtained or derived from sources believed to be reliable. Neither the authors nor J.P. Morgan makes any representations or warranties as to the information's accuracy or completeness. The information contained herein has been provided solely for informational purposes and does not constitute an offer, solicitation, advice or recommendation, to make any investment decisions or purchase any financial instruments, and may not be construed as such. All expressions of opinion in the paper are subject to change without notice and reflect the judgment of the authors as of the date of publication. No responsibility is taken for changes in market conditions or laws or regulations and no obligation is assumed to revise this paper to reflect changes, events or conditions, which occur subsequent to the date hereof. Similarly, the material contained in this paper does not constitute, and should not be relied on as, legal, regulatory, accounting, tax, investment, trading or other advice. Any financial, tax, or legal information contained in this paper was so included for informational purposes only. You should consult your own financial adviser, tax adviser or legal counsel, as appropriate. Neither the authors nor J.P. Morgan is responsible for any error, omission or for the interpretation of any law or regulation. The opinions and recommendations herein do not take into account individual circumstances, objectives, or needs and are not intended as recommendations for the sale or purchase of particular securities, financial instruments or strategies. The recipient of this paper must make its own independent decisions regarding the information mentioned herein. This paper does not bind the authors or J.P. Morgan in any way.

Additional disclosures and other information are available at: www.jpmorgan.com/pages/disclosures.

JPMorgan Chase Bank, N.A. Member FDIC. Deposits held in non-U.S. branches are not FDIC insured.

JPMorgan Chase Bank, N.A., organized under the laws of U.S.A. with limited liability.

© 2026 JPMorgan Chase & Co. All Rights Reserved