

The financial infrastructure for AI agents: a 2026 field report

What has actually been built for agents to pay, hold, and be trusted with money as of September 2026. Rails, protocols, wallets, controls, and credit, with the industry's own numbers and where they disagree.

Nakul Kelkar Co-founder, Vaaya

[X](#) [LinkedIn](#)

A PDF of this report is at vaaya.ai/reports/financial-infrastructure-for-ai-agents-2026.pdf.

This is a report on what exists, not on what is coming. Eighteen months ago the financial infrastructure for AI agents was a whitepaper, a reserved HTTP status code, and a great deal of speculation. As of September 2026 it is a set of live rails processing tens of millions of machine-initiated payments a month, two competing families of protocol with the largest card networks, processors, and model labs signed on to one or both, a first generation of wallets and gateways, and the beginnings of a policy and credit layer. This document lays out each piece with the industry's own figures, notes where those figures disagree, and is candid about the layers that are still missing. Where a number is given, its source and date are given with it. The two of us run a company that builds one of the pieces described here, and it is named where relevant; the numbers are not ours.

Summary

Machine-to-machine payment rails are real and in production, and their volume is measured in transactions rather than dollars: by Coinbase's count, x402 had cleared 165 million payments and about \$50 million in cumulative value by April 2026, which is an average of thirty cents or less per payment. Stripe and Tempo launched the Machine Payments Protocol in March 2026 with a directory of over a hundred compatible services and a set of design partners that includes both card networks and both leading model labs. Stablecoins are the settlement asset for nearly all of it: USDC accounted for over 99.99 percent of agentic transfer volume on x402 over the ninety days to mid-2026, and USDC's own on-chain volume reached \$21.5 trillion in the first quarter of 2026. The merchant-checkout family of protocols, ACP, AP2, and UCP, is live in ChatGPT and Google's surfaces but still early, and the card networks have completed their first hundreds of live agent-initiated transactions rather than their first millions.

Forecasts for agentic commerce by 2030 range from Bain's \$300 to \$500 billion for the United States alone to McKinsey's \$3 to \$5 trillion globally, a spread wide enough that the honest reading is that nobody knows the size, only the direction. What has been built is settlement. What has not been built, outside a handful of gateways, is identity, authorization, a record that accumulates, and credit. Those are the layers that turn payments into a market, and they are where the next two years will be decided.

1. Definitions

Agentic payments are payments initiated by software on its own decision, without a person approving each one. The paradigm case is an agent paying an API for a search, a scrape, a minute of compute, or a row of data, per call, in amounts from a fraction of a cent to a few dollars.

Agentic commerce is an agent buying from a merchant on a person's behalf: a flight, a pair of shoes, groceries. It is settled on cards, carries consumer-protection expectations, and needs a proof that the person authorized the agent. It is a different problem from agentic payments and shares a word with it, which causes most of the confusion in this field.

Agentic finance infrastructure is the set of systems that let software hold, spend, receive, and be extended money. It has four layers, and the layers are the organizing principle of this report.

1. **Identity.** Whose money is this, and which agent is spending it. Without a persistent identity per agent there is no attribution, no history, no revocation of one agent without the others, and later, nothing to underwrite.
2. **Authorization.** What this agent may buy, from whom, and up to how much, decided before the payment and enforced by something the agent cannot argue with.
3. **Settlement.** The rail value moves on. The rail sets the fee floor, which sets the smallest economic purchase, which decides whether an agent can meter its spending or must buy in bundles like a person.
4. **Record.** What was bought, from whom, at what price, on which key, and what was refused and why. The record is what makes the other three auditable, and it is the raw material for reputation and credit.

2. Settlement: the rails

2.1 Stablecoins are the asset

Every serious machine-to-machine rail settles in a dollar stablecoin, and in practice in USDC. Circle reported USDC in circulation of \$77.0 billion at the end of the first quarter of 2026, up 28 percent year on year, and on-chain transaction volume for the quarter of \$21.5 trillion, up 263 percent. Visa's adjusted on-chain analytics, which strip out exchange transfers and bot activity, put total stablecoin transaction volume at a record \$1.79 trillion for June 2026, with USDC at roughly 67 percent and USDT at roughly 32 percent, and USDC at about 70 percent of adjusted volume across the first half of the year. Total stablecoin market capitalization passed \$315 billion in the same period.

The relevance to agents is the fee floor. The x402 whitepaper, published by Coinbase Developer Platform in May 2025, framed the problem directly: card fees "as high as \$0.30 per transaction" make microtransactions impractical, while on-chain settlement completes in roughly 200 milliseconds at a cost that makes one-cent payments viable. On Base, Coinbase's Layer 2, a USDC transfer costs a fraction of a cent. That number is the reason agents can meter and people never could.

2.2 x402: pay per request, on-chain

x402 is an open standard, Apache-2.0 licensed and maintained by the x402 Foundation, that revives HTTP status 402. A server answers an unpaid request with 402 and a `PAYMENT-REQUIRED` header carrying the amount, asset, network, and address; the client signs an authorization and retries with a `PAYMENT-SIGNATURE` header; a facilitator with `/verify` and `/settle` endpoints confirms and executes the transfer; the server returns the resource. Base and Solana are supported, any ERC-20 or SPL token is accepted in principle, the buyer pays gas, and most facilitators charge no fee of their own. Settlement takes about a second.

The volume figures, in order of their source's proximity to the data:

Figure	As of	Source
165 million transactions, about \$50 million cumulative volume, roughly 69,000 active agents	April 21, 2026	Coinbase, at the launch of its agent marketplace
"More than 100 million x402 payments across Base and Solana" processed by the CDP facilitator	Current documentation	Coinbase Developer Platform docs
120 million or more transactions, \$41 million or more settled, 14 chains, 8 active facilitators, \$0.05 average payment; Base about 58 percent of volume, Solana about 38 percent	April 14, 2026	web3trackers.com x402 dashboard
Cumulative transactions past 160 million, more than 90 percent on Base	June 2026	Crypto Briefing
14 to 17.8 million transactions in a 30-day window	August 2026	Crypto Briefing
USDC over 99.99 percent of agentic transfer volume on x402 over 90 days; average transaction on Base roughly \$0.13	Mid-2026	Crypto Briefing, citing Token Terminal

The figures disagree in their scope, not in their story. Trackers count different chains and facilitators; Coinbase counts what its own facilitator saw. Read together: on the order of 150 to 200 million payments have cleared over x402 in about a year of real use, the great majority on Base, at an average value between five and thirty cents depending on the window, with tens of millions of payments a month by the summer of 2026. In dollar terms the whole thing is smaller than one large merchant's day. In count it is already more payments than most card programs process, and that is the correct lens: agents make many small decisions, and the rail exists to make each one cheap.

2.3 MPP: pay per request, through a processor

The Machine Payments Protocol was co-authored by Stripe and Tempo and released in March 2026 alongside the mainnet launch of Tempo, the payments-focused Layer 1 that Stripe and Paradigm incubated. The flow is the same shape as x402: an unpaid request draws a 402 challenge, the agent retries with a payment credential, the server records the payment and returns the resource with a receipt. What differs is what settles. An agent can pay with a stablecoin sent to a deposit address on Tempo, which Stripe off-ramps into the seller's

balance, or with a card through a shared payment token provisioned via Link, in which case Stripe's fifty-cent card minimum applies. Tempo settles gas in stablecoins rather than a native token. MPP also introduced a sessions primitive that lets an agent authorize a spending limit up front and then stream micropayments against it without an on-chain transaction per interaction, which is the first protocol-level acknowledgement that agents need a limit, not just a payment.

At launch MPP shipped a payments directory of over a hundred compatible services, with Stripe providing card methods, Visa wallet methods, and Lightspark Bitcoin Lightning. Tempo's design partners as reported at launch included Visa, Mastercard, Deutsche Bank, Standard Chartered, Revolut, Nubank, Shopify, OpenAI, Anthropic, Ramp, and DoorDash. No transaction volume has been disclosed for MPP or Tempo as of this writing.

The two rails are not rivals so much as the same idea landing in two ecosystems. A provider that already runs on Stripe adopts MPP; a provider that is chain-native adopts x402; a buyer that wants both needs something that speaks both.

2.4 Cards, extended toward agents

The card networks have not built a new rail. They have built consent and tokenization schemes so that an agent can transact on an existing card with the cardholder's delegated authority. Visa announced on December 18, 2025 that it and its partners had completed hundreds of secure agent-initiated transactions in live production, that more than a hundred partners were working with Visa Intelligent Commerce, more than thirty were building in its sandbox, and more than twenty agents and agent enablers were integrating directly, and predicted that millions of consumers would use agents to complete purchases by the 2026 holiday season. Mastercard's Agent Pay binds a tokenized card credential to a specific agent, merchant scope, and consent policy through what it calls Agentic Tokens; in March 2026 it completed live agentic transactions across Latin America with seventeen participating processors and issuers, and with Santander completed what the two described as Europe's first live end-to-end payment executed by an AI agent inside a regulated banking framework. On August 31, 2026, Visa, Mastercard, and Fiserv joined a newly formed Agentic Payments Alliance to set shared standards for agent identity, loyalty, and fraud detection.

These are milestones measured in hundreds of transactions and dozens of partners. They matter because they are the path by which agentic commerce, the merchant-checkout kind, reaches the consumer, and they carry card economics with them: a fixed fee per transaction

that makes a two-cent payment absurd. Cards will settle the flight. They will not settle the search.

3. Protocols: how the exchange is structured

Five protocols have a serious claim to the phrase, and they belong to two families.

	x402	MPP	ACP	AP2	UCP
Who	x402 Foundation; Coinbase runs the largest facilitator	Stripe and Tempo	OpenAI and Stripe	Google, donated to the FIDO Alliance	Google
Family	Pay per request	Pay per request	Merchant checkout	Merchant checkout	Merchant checkout
What settles	USDC or any token on Base or Solana	Stablecoin on Tempo, or a card via shared payment token	The merchant's processor: Stripe, Adyen, Braintree	Cards today; wallets, bank transfers, digital currencies on the roadmap	Google Pay
Human in the loop	No	No	Yes, in ChatGPT	Yes, via signed mandates	Yes
Status	Production, tens of millions of payments a month	Production since March 2026; volume undisclosed	Live in ChatGPT; discovery still being built	v0.2; standardization at FIDO	Evolving; waitlisted

The pay-per-request family solves metering: how a piece of software pays another for one unit of work without an account or a checkout. The merchant-checkout family solves delegation: how a merchant and a card network can trust that a person authorized this agent to make this purchase. ACP structures a checkout session inside ChatGPT with a delegated payment token so the agent never holds the card. AP2 encodes the person's authorization as cryptographically signed mandates, a checkout mandate and a payment mandate, that create

a non-repudiable audit trail; Google donated it to the FIDO Alliance, where standardization is under way. UCP is the protocol merchants implement to sell inside Google Search and Gemini. Almost nothing built for one family is useful for the other, and the single most common error in this field is to compare across the line.

4. Identity and custody: wallets

A rail moves value. Something has to hold it, and there are three designs.

The agent holds a key. Coinbase's AgentKit and Circle's Agent Wallets, announced with Circle's Agent Stack in the first quarter of 2026, give an agent an on-chain wallet it controls, and ship policy engines alongside because a funded key with no policy is a liability. This is the right design when the agent is itself a market participant that pays other agents or holds value for its own reasons. Coinbase counted roughly 69,000 active agents on x402 in April 2026, and each of those is, in some form, this design.

The agent holds a card. Stripe Issuing and the products built on it for agents provision a virtual card per agent with issuer controls on amount and merchant. This is the right design for merchant checkout and the wrong one for APIs, because it imports the card fee floor.

The agent holds permission. A person or company holds the account; each agent holds a key with a policy; a gateway holds the provider relationships and pays each one on whatever rail it accepts. The agent never touches money. This is the design for an agent that is an employee rather than a participant, which is most agents, and it is the one Vaaya is built on.

The custody choice decides liability. An agent with a funded key and a bug can drain the key. An agent with a key to a policy and the same bug hits a ceiling and stops. Both are called wallets, and the difference between them is the difference in what a mistake costs.

5. Authorization: what a policy layer has to enforce

The rails have almost nothing to say about how much an agent may spend; MPP's sessions primitive is the first exception, and it is a limit on one session with one seller. Everything else has to be enforced by whatever stands between the agent and the rail. From the deployments we have seen and built, a policy layer for an agent has to provide at least the following, and has to provide them before the payment rather than on the statement.

1. **A maximum cost per call, stated by the agent.** The agent has to say what it will pay for this unit of work, every time, and the system has to refuse anything above it. This is the control that bounds the single expensive mistake and makes price a decision rather than a discovery.
2. **A limit per agent per period**, attached to the agent's own key so that one agent's leak does not drain the account and one agent can be revoked without the others.
3. **A scope**: which categories of thing the agent may buy at all. A research agent that cannot start a sandbox is a research agent that cannot be talked into starting one.
4. **A ceiling on the whole account**, independent of how many keys exist.
5. **A human on the irreversible action only.** Approval on every five-cent call becomes a rubber stamp by the fifth call. Approval on a sent email, which cannot be refunded, is the one place a person belongs.
6. **Refusals that cost nothing and are recorded.** A call above a limit should never reach the provider, should never be billed, and should leave a record with its reason. The refusal log is the most informative artifact the layer produces.

None of this is exotic. All of it is absent from a bare rail, and most of it is absent from most agent frameworks, which is why the OWASP entry for excessive agency exists.

6. Credit: the missing layer

Every rail and wallet above assumes the money is already there. That assumption is what limits the market. A prepaid agent spends defensively, buys only from providers it already knows, and produces no information about its principal beyond the fact that they had the funds. Credit is different in kind, not degree. **Credit is not a liquidity product. Credit is an information production mechanism that is sold as liquidity.** To extend it, someone has to assign the agent a persistent identity, attach it to a principal, decide what that principal is good for, and then watch every draw and every settlement. Each of those acts produces a record that did not exist before, and the record is what a market needs to scale beyond the participants who already trust each other.

Agents cannot be underwritten directly in 2026. They have no history that survives a redeploy and no recourse. Credit has to attach to the person or company behind the agent, and the interesting question is what a lender can know about that person that a bank cannot. For

developers the answer is their public work: years of commits, reviews, and maintained repositories, timestamped and hard to fake at scale. Vaaya sizes a credit line from that record, lets the developer's agents draw on it from the first day inside a policy, and settles it weekly against a card. The line is deliberately small to begin with and grows with reliable settlement. We do not publish its sizes or our volumes, and this report does not rely on them. The mechanism is the point: it lets an agent spend before value has been deposited, and it starts the agent's own economic history from its first call. In a few months of high-frequency transactions an agent accumulates the kind of record a person accumulates in years, and that record is what will eventually let the agent itself be the thing underwritten.

7. What the forecasts say, and why they disagree

Source	Forecast	Date
Juniper Research	About \$8 billion of agentic commerce spend in 2026, \$1.5 trillion by 2030, \$3.5 trillion by 2031	April 7, 2026
McKinsey	\$3 to \$5 trillion globally by 2030	2025
Bain	\$300 to \$500 billion in the United States by 2030, 15 to 25 percent of e-commerce	2025
Morgan Stanley	10 to 20 percent of US e-commerce sales agent-driven by 2030	2025
Gartner	20 percent of digital commerce transactions executed through AI platforms by 2030	2025
Edgar Dunn	\$136 billion in 2025 to \$1.7 trillion by 2030, a 67 percent compound rate	2025
Adobe Analytics	AI-driven visits to US retail sites up 4,700 percent year on year in 2025	2025

The spread between Juniper's \$8 billion for this year and McKinsey's trillions for 2030 is not a disagreement about growth. It is a disagreement about what is being counted. The consumer forecasts count retail purchases that an agent touched, which is agentic commerce and will be settled on cards. None of them count the machine-to-machine market at all, because in dollars it is small, and a \$0.13 payment does not move a trillion-dollar forecast. The correct way to

hold both is that the consumer side will be large in dollars and slow to arrive, because it needs consent, dispute, and regulation to be solved first, and the machine side is already large in count and will stay small in dollars per payment forever, because that is the point of it. The infrastructure this report describes is mostly for the second.

8. What is not built

Recourse. No machine-to-machine rail has dispute resolution. x402 settlement is final in about a second, which is a feature for sellers and a gap for buyers. If a provider returns a poor result, the payment stands. The merchant-checkout protocols inherit card disputes; the pay-per-request protocols inherit nothing.

Reputation. There is no shared rating for sellers on either rail, and no shared history for agents. Each gateway keeps its own. The IMF's April 2026 note on how agentic AI will reshape payments and Visa's agentic-payments research both identify trust as the binding constraint, and trust is a record problem before it is a technology problem.

Agent-to-agent trust. Two agents transacting directly have signatures and nothing else. The mandate model in AP2 is the most developed answer for the consumer case and does not apply to the machine case.

Merchant checkout outside the networks. An agent with a policy and a gateway can pay for anything that is an API. It cannot buy a flight. Cards can, and the consent layer that makes that safe is still in pilots measured in hundreds of transactions.

Credit at scale. Underwriting attaches to the principal today, and the principals who can be underwritten are those with public records. Companies with KYB and individuals with KYC are the next step, and the data to underwrite the agent itself is still accumulating.

9. Conclusion

The financial infrastructure for AI agents in September 2026 is a settlement layer that works, two protocol families that have each attracted the incumbents they needed, a wallet layer that is mature for agents that hold keys and early for agents that hold permission, a policy layer that exists only inside gateways, and a credit layer that is just beginning. The rails were the hard part and they are done; they are also the part that will be commoditized. What is not commoditized, and cannot be, is the record: the per-agent history of what was bought, at what

price, and what was refused, because it accumulates and everything else in the stack does not. Payment infrastructure that keeps that record from the first call is building the asset that reputation, insurance, routing, and credit will all be built on. Payment infrastructure that only moves value is a utility, and utilities are priced accordingly.

Sources

- Coinbase Developer Platform, x402 whitepaper, "x402: An open standard for internet-native payments," May 6, 2025.
- Coinbase Developer Platform, x402 documentation, "Welcome to x402," current.
- x402 Foundation, docs.x402.org, introduction and FAQ, current.
- Coinbase, agent marketplace launch figures as reported April 21, 2026: 165 million transactions, about \$50 million cumulative volume, roughly 69,000 active agents.
- web3trackers.com, x402 dashboard, as of April 14, 2026.
- Crypto Briefing, "USD Coin dominates agentic transfer volume," and related x402 coverage, mid-2026.
- Circle, "Circle Reports First Quarter 2026 Results," and the Agent Stack announcement.
- Visa Onchain Analytics via Allium, June 2026 stablecoin volume, as reported by Crypto Briefing.
- Stripe, Machine Payments Protocol documentation, docs.stripe.com/payments/machine/mpp, current.
- Ledger Insights and The Defiant, Tempo mainnet and MPP launch coverage, March 2026.
- Visa, "Visa and Partners Complete Secure AI Transactions," December 18, 2025.
- Mastercard, Latin America and Santander agentic transaction announcements, March 2026.
- The Motley Fool, Agentic Payments Alliance coverage, August 31, 2026.
- Google, AP2 documentation, ap2-protocol.org; Google for Developers, UCP overview.
- OpenAI and Stripe, Agentic Commerce Protocol, agenticcommerce.dev.
- Juniper Research, "Agentic Commerce Market Report," April 7, 2026.
- McKinsey, Bain, Morgan Stanley, Gartner, Edgar Dunn, and Adobe Analytics forecasts as compiled by commercetools, 2026.

- International Monetary Fund, "How Agentic AI Will Reshape Payments," IMF Note, April 22, 2026.
- OWASP GenAI Security Project, LLM06:2025 Excessive Agency.

Questions

What is agentic finance infrastructure?

The systems that let software hold, spend, receive, and be extended money without a person initiating each act. It has four layers: identity (whose money, which agent), authorization (what it may buy, up to how much), settlement (the rail value moves on), and record (what was bought, at what price, and what was refused). Rails are the most built layer; record and credit are the least.

What are the layers of AI agent payment infrastructure?

Identity, authorization, settlement, and record. Settlement is provided by stablecoin rails and card networks, structured by protocols such as x402, MPP, ACP, AP2, and UCP. Identity and authorization are provided by wallets and gateways. The record is what accumulates into reputation and, eventually, credit.

How large is agentic commerce in 2026?

Small in dollars and large in count. Juniper Research put agentic commerce spend at about \$8 billion for 2026, forecasting \$1.5 trillion by 2030. On the machine-to-machine side, Coinbase reported 165 million x402 transactions and roughly \$50 million of cumulative volume by April 2026, at an average of well under a dollar per payment.