What Is ERC-8004?

A draft Ethereum standard that gives AI agents on-chain identity, reputation, and verifiable trust.

Alain AI Lab Research · Published July 14, 2026 · 10 min read

ERC-8004 is a proposed Ethereum standard, informally titled “Trustless Agents,” that gives autonomous AI agents a way to prove who they are and whether they can be trusted — without any human vouching for them. It defines shared on-chain registries where an agent can hold a portable identity, accumulate a reputation, and have its work independently checked. The point is narrow but important: as software agents begin to transact with one another over open networks, they need a common answer to the oldest question in commerce — can I trust the party on the other side? A standard supplies that answer in a form every agent can read. For the wider picture of how these agents operate on-chain, our report on what an AI agent in crypto is sets the stage.

AT A GLANCE

WHAT IT IS **Draft Ethereum standard for AI-agent trust**	NICKNAME **“Trustless Agents”**	CORE PIECES **Identity, reputation, validation registries**
STATUS **Early-stage proposal, not final**	SITS BESIDE **A2A, x402 payments, ERC-4337 wallets**	SOLVES **Trust between agents that never met**

01

What ERC-8004 Actually Is

Strip away the branding and ERC-8004 is a set of rules for three shared record-keeping contracts that live on a public blockchain. The first lets an AI agent register a durable identity — an on-chain reference that stays the same no matter which platform the agent is running on. The second records feedback about that agent, so past behaviour becomes a

readable history rather than a marketing claim. The third provides a way for third parties to attest that the agent actually did what it says it did. Together they aim to turn “trust me” into “check the record.”

The word *trustless* in the nickname is doing precise work. It does not mean no trust is required; it means trust does not have to rest on a private relationship or a central gatekeeper. Instead it rests on shared, verifiable state that anyone — human or machine — can inspect. That is the same principle that underpins most of crypto: replace “take my word” with “read the ledger.” ERC-8004 simply applies it to the identity and reputation of software agents rather than to the movement of tokens.

02

Why a Big Number Means a New Idea

The name looks cryptic but follows a simple convention. ERC stands for Ethereum Request for Comments, a category of Ethereum Improvement Proposal — an EIP — used for application-level standards that developers can build against. ERC-20 defined fungible tokens; ERC-721 defined non-fungible tokens; ERC-4337 defined smart-contract wallets. Each number is assigned roughly in the order proposals arrive, so a high figure like 8004 is a strong signal that this is a recent arrival, not a settled pillar of the ecosystem.

That framing matters for expectations. A low-numbered standard like ERC-20 is battle-tested by years of use and billions of dollars in value. A four-digit newcomer is closer to a well-argued draft: published, debated, and implemented in demos, but still short of the broad adoption that turns a proposal into infrastructure. Reading ERC-8004 as an early blueprint rather than a finished foundation keeps the analysis honest.

03

The Problem: Trust Between Strangers

To see why the standard exists, picture two AI agents that have never interacted. One needs a task done — summarise a dataset, price a risk, execute a swap — and the other advertises that it can do it. In the human world we resolve this with reputations, references, brands, and courts. Agents on an open network have none of that by default. They meet as strangers, and a stranger with a payment address is indistinguishable from a scammer with the same.

The problem sharpens because agents are meant to act without a human approving each step. If a person had to vet every counterparty, the autonomy that makes agents useful

would evaporate. So the trust decision has to be something an agent can make on its own, quickly, from evidence it can verify. ERC-8004 is an attempt to supply exactly that evidence — a machine-readable answer to "is this agent real, and has it behaved well before?" — drawn from a source neither party controls.

> A payment rail lets one agent pay another. It says nothing about whether the other agent is competent or honest. ERC-8004 is the missing half: the trust layer that decides whether the payment should happen at all.

04

The Three Registries

The proposal is usually described in terms of three cooperating registries, and the division is the clearest way to understand it. The **Identity Registry** is the anchor. It gives each agent a persistent on-chain identifier and a place to point at its own description — an "agent card" of capabilities and endpoints — so that an identity can be discovered and referenced consistently rather than reinvented on every platform.

The **Reputation Registry** is the memory. It provides a structured way to attach feedback to an agent's identity, so that completed jobs, ratings, or attestations become part of a history other agents can read before deciding to engage. The **Validation Registry** is the auditor. It offers a mechanism for independent parties to confirm that a piece of work was genuinely performed — through re-execution, staking, or cryptographic attestation — rather than merely claimed. Identity says who; reputation says how they have behaved; validation says whether a specific claim checks out.

Importantly, these registries are meant to be minimal and composable. They do not try to score agents or dictate a trust policy. They store verifiable facts and leave the judgement — how much reputation is enough, which validators to believe — to whoever is using the data. That restraint is deliberate: a standard that hard-codes opinions ages badly, while one that stores clean primitives lets many trust models be built on top.

05

How It Fits the Agent Stack

ERC-8004 is not designed to stand alone; it is one layer in an emerging stack for agent-to-agent commerce. Above it sit communication protocols — most visibly the Agent-to-Agent

(A2A) approach popularised in the Google ecosystem — which define how agents describe themselves and exchange requests. ERC-8004 borrows the "agent card" idea from that world and anchors it on-chain, so identity survives beyond any single platform's servers.

Beside it sit the payment rails. The x402 protocol, associated with work at Coinbase, revives the dormant HTTP "402 Payment Required" status so that a machine can pay for a resource inline. And beneath both sit the wallets that let an agent hold and move value safely: ERC-4337 account abstraction supplies smart-contract wallets with features like session keys, which grant an agent narrow, time-boxed spending permission instead of an unrestricted private key. In that division of labour, communication protocols carry the conversation, x402 carries the money, ERC-4337 holds the funds, and ERC-8004 answers whether the counterparty deserves any of it.

06

The Agent Economy It Points Toward

The reason this attracts attention is the scenario it enables: a machine-to-machine economy in which agents hire other agents. An orchestrating agent could discover a specialist it has never used, read its on-chain reputation, commission a task, verify the result through a validator, and pay — all without a human in the loop. That is a natural extension of the coordination patterns explored in our report on what a multi-agent system is, pushed out of one organisation and onto an open market.

For that market to function, the trust decision has to be portable and cheap. A specialist agent should be able to build a reputation once and carry it everywhere, and a buyer agent should be able to price the risk of a stranger from public evidence. ERC-8004 is a bet that this coordination problem is best solved at the standard layer — as shared infrastructure — rather than re-solved privately inside every platform. Whether that bet pays off depends less on the elegance of the design than on how many builders actually adopt it.

07

Where It's Real and Where It Isn't

Honesty about maturity matters here. ERC-8004 is an early-stage proposal that surfaced as the agent-payments conversation accelerated through 2025. It has reference implementations, demonstrations, and active discussion, and it draws on credible work at the intersection of the Ethereum research community and the agent-interoperability efforts.

What it does not yet have is the thing that makes a standard real: widespread, load-bearing use. A specification existing is not the same as an ecosystem relying on it.

So the accurate reading is "promising and plausibly important, but unproven." The design addresses a genuine gap, the surrounding stack is being built in parallel, and the timing lines up with real demand for autonomous agents that transact. But numbers on adoption, the final shape of the registries, and whether the broader industry converges on this standard rather than a rival are all open. Treating ERC-8004 as a settled foundation would be premature; treating it as a serious proposal worth understanding is exactly right.

08

The Hard Parts: Sybil, Privacy, Permanence

Putting identity and reputation on a public chain solves some problems and creates others. The sharpest is the **Sybil problem**: if creating an identity is cheap, a bad actor can spin up many agents, have them praise each other, and manufacture a reputation from nothing. Any on-chain reputation system has to make honest history costly to fake, usually by tying it to stake, real completed work, or attestations from validators that themselves have something to lose — and getting that economic design right is unsolved, not automatic.

The second tension is **privacy versus permanence**. A blockchain's strength is that records are durable and public; its weakness, for identity, is exactly the same. A mistaken piece of negative feedback, or data an agent's operator would rather not expose, is hard to erase once written. And every registry an agent touches is another surface where a hostile instruction or a malformed attestation could cause harm — concerns explored in our report on AI agents in DeFi and their security breaches. None of this makes the idea unworkable. It means the useful question is not "is ERC-8004 clever?" but "does its trust hold up when someone is actively trying to game it?" That test, more than any specification detail, will decide whether trustless agents become real infrastructure or stay an elegant draft.

"A good name is rather to be chosen than great riches."

PROVERBS 22:1

METHODOLOGY & SOURCES

This report describes ERC-8004 as an early-stage, proposed Ethereum standard and deliberately avoids stating adoption figures, precise authorship, or exact dates that are not firmly established in

public. Where the standards process is still moving, claims are framed as approximate and provisional; no specific decimals, dollar figures, or user counts were invented, and the draft status of the proposal is stated throughout rather than glossed over.

Primary framing draws on the Ethereum Improvement Proposal process and the ERC category (as used by ERC-20, ERC-721, and ERC-4337 account abstraction); the “Trustless Agents” proposal’s three-registry design for identity, reputation, and validation; the Agent-to-Agent (A2A) interoperability approach and its “agent card” concept; the x402 inline-payment protocol associated with Coinbase; and standard security literature on Sybil resistance and on-chain data permanence. Readers should consult the live EIP repository for the authoritative, current text. Educational content only; not financial advice.

What Is an AI Agent in Crypto? · What Is a Multi-Agent System? · AI Agents in DeFi: Security Breaches

Alain AI Lab — independent crypto & AI research.
intelligencecrypto.org · This document is for educational purposes only and is not financial advice.