# Identity, Trust, and the Legal Foundations of Agentic Commerce

**A Proposal for Trust in Depth**

David Fisher, CEO, Integra Ledger
Bridget McCormack, CEO, American Arbitration Association

April 2026

## I. The Problem

The agentic commerce era is upon us. Hundreds of the world's largest companies have lined up behind new protocols for agentic payments, and transaction volumes are growing exponentially. Gartner projects $15 trillion in AI-intermediated B2B commerce by 2028. But something important is missing - the trust infrastructure that has anchored free markets for centuries.

This gap creates significant risk. In 2024, automated bot traffic surpassed human activity on the internet for the first time: 51% of all global web traffic is now non-human (Imperva, 2025). Researchers at ETH Zurich demonstrated a 100% solve rate against Google's reCAPTCHA using fine-tuned AI models (Plesner et al., 2024). The gate that separates humans from machines is gone. Every platform identity system, every electronic signature, every click-through agreement, and every CAPTCHA was built on an implicit assumption: the actor on the other side is a person, operating at human speed, constrained by human limitations, and governed by human-built guardrails. AI agents disrupt all of this.

The pre-agentic world of e-commerce operated within an acceptable trust equilibrium. Friction was a design spec of that equilibrium. The cumbersome processes of presenting identification, waiting for verification, and clicking through terms of service were the trust mechanisms, and the legal infrastructure supported them. The cost of gaming them was high enough to deter most bad actors.

AI agents eliminate that economic deterrent. They can create synthetic identities at scale, pass verification checks with generated documents, accept terms without any intention of honoring them, and establish reputation through manufactured history. Each gap that previously required expensive human effort to exploit can now be exploited programmatically, instantly, and at massive scale.

Law is the operating system for civilization. It organizes how humans relate to each other, how they establish trust, and how they resolve disagreements. These patterns are distilled into legal systems that enable social order to scale beyond families, tribes, and villages. Without them, commerce can't function, property rights are unenforceable, and there is no

governance around agreements. AI agents are the first autonomous actors to arrive in a world with no framework for connecting their actions to our legal operating system.

The industries supporting agentic commerce have been focused on payments. Agents with funded cryptocurrency wallets can initiate transactions, settle invoices, and move value across borders in seconds. Stablecoins, DeFi protocols, and machine-to-machine payment rails have attracted billions in investment. But commerce is more than payment. It is the full lifecycle of economic relationships — negotiation, agreement, performance, and dispute resolution. A system that lets agents pay but provides no mechanism for resolving problems when payments go wrong, services fall short, or obligations aren't met is fundamentally incomplete.

A working payment layer is necessary but not sufficient. Agentic commerce also needs a working identity layer, agreement layer, and dispute resolution layer. This paper proposes a framework for what's missing: trust in depth.

## II. Code (Alone) Is Not Law

The historical evolution of trust in commerce is important background framing for understanding why agentic commerce introduces a fundamental break from that history.

### Progression

The history of trust in commerce is a story of identity progressively detaching from the individual.

The signature was the first modern technology for binding a person to a commitment. Its authority rested on a social contract, not a technology: the signer's unique mark carried the signer's commitment, and the legal system would enforce it. Identity and individual were the same.

Electronic signatures shifted the trust anchor from the individual to a platform. A login to Docusign creates an identity with Docusign, and Docusign becomes the witness, the notary, and the custodian of the record. As Amber Baldet, then executive director of JPMorgan's Blockchain Center of Excellence, observed at the launch of the Enterprise Ethereum Alliance in February 2017, we may have computerized previously analog and social systems, but we didn't digitize them. E-signatures computerized the signing process. They also quietly centralized the trust anchor. That is the model the entire cloud era runs on. But the model breaks when you remove the intermediary.

Public Key Infrastructure attempted the right fix—cryptographic binding of identity to consent. It failed because the implementation was too complex for everyday use. PKI found its home in TLS/SSL and enterprise authentication, the invisible infrastructure that secures all digital communication, but never replaced the wet signature for routine commerce. The core concept was right, usability was not.

Blockchain proposed a radical inversion: instead of routing every interaction through a trusted middleman, two parties could transact directly, and the protocol provided the trust. This was a genuinely revolutionary shift, and the first technology in thousands of years to reverse the progressive centralization of trust anchors. Blockchain is extraordinarily good at proving that something happened: this transaction occurred, at this time, signed by this key, immutably. No other technology has offered universally available public proof of private transactions without a central verification authority.

But blockchain cannot bridge from cryptographic proof to real-world identity. It proves that key `0x7f3a...` signed a transaction. It cannot prove that key `0x7f3a...` belongs to Jane Smith, CFO of Acme Corp, incorporated in Delaware and subject to New York law.

## The Promise of Smart Contracts

Smart contracts—the execution layer built on blockchain—are central to the architecture of agentic commerce. But they are widely misunderstood, in large part because the term implies capabilities the technology does not possess. Understanding what smart contracts were meant to be, and what they became, is essential context for the framework this paper proposes.

Nick Szabo, a computer scientist and legal scholar coined the term in 1994 and developed it substantially in a 1996 paper, "Smart Contracts: Building Blocks for Digital Markets." Szabo's vision was grounded in computer science and in contract law. As Szabo explained, the contract, "a set of promises agreed to in a 'meeting of the minds,'" is the foundational building block of a free market economy, and common law principles of contract represent centuries of cultural evolution that would be "prohibitively costly to recompute." His view was that these principles should not be replaced, but rather extended: "What parts of our hard-won legal tradition will still be valuable in the cyberspace era?"

His answer was ambitious. Szabo defined a smart contract as "a set of promises, specified in digital form, including protocols within which the parties perform on these promises." The language was deliberately contractual: promises, performance, parties. He identified four design principles drawn directly from common law contract and economic theory: observability (parties can monitor performance), verifiability (a third party can determine whether breach occurred), privity (knowledge and control distributed only as necessary for performance of the contract), and enforceability (minimizing the need for external enforcement while preserving it as a backstop). Critically, his architecture included arbitrators as a core feature to be "invoked to resolve disputes arising out of performance (or lack thereof.)" Szabo's project was to combine the strengths of legal and digital infrastructure.

The blockchains that were founded many years later could not have delivered on that full vision. A trustless blockchain requires absolute determinism; every node must produce the same result, every time, with no room for interpretation. It is this constraint that makes it trustworthy. It also makes Szabo's broader aspiration extraordinarily difficult to realize, because the legal principles he was trying to embed are not exclusively deterministic.

Contracts depend on judgment, context, and principles of reasonableness and good faith that require interpretation, precisely the capacity that deterministic systems, by design, do not possess. Szabo acknowledged this.

The result was a category error that persists today. The term "smart contract" became ubiquitous, carrying an implied association with legal contracts that the technology could not deliver. In practice, smart contracts are rigorous, deterministic code, extraordinarily reliable at enforcing "if X then Y." They can transfer cryptocurrency, swap tokens, execute a buy or sell. But they are not contracts in any sense a lawyer, a court, or a counterparty would recognize.

A transaction is a moment. An agreement is a relationship that extends in time.

Contracts involve negotiation, interpretation, good faith, remedies, and evolving circumstances. The Romans introduced *bona fides* over two thousand years ago to give legal heft to the idea that parties must deal honestly beyond the literal terms. It remains foundational to contract law in virtually every modern jurisdiction. Contracts require understanding of intent, context, and ambiguity, and their enforceability is governed by principles such as "reasonableness," "materiality," and "best efforts." Even a twenty-dollar purchase carries the invisible protection of consumer protection law, product liability doctrine, and implied warranties. Deterministic code has no access to any of this.

The assertion that "code is law" completed the drift, collapsing the distance between execution and agreement entirely and claiming that running a program *is* the legal relationship. Code can be an extraordinarily reliable execution layer. But execution is not agreement. Agreement requires identity, consent, terms, jurisdiction, and recourse, none of which code provides.

AI is the first technology with the plausible capacity to operate in this interpretive space — to supply the judgment, context, and reasoning about intent that Szabo's vision required but deterministic systems could not provide. The combination of blockchain's trusted execution with AI's capacity for interpretation points, for the first time, toward the possibility of realizing smart contracts as Szabo originally conceived them: automated systems that can engage with the richness of contractual relationships, not merely execute transactions.

## Severance

But that promise surfaces a new and fundamental challenge. The very autonomy that makes AI capable of operating in the interpretive space of contracts also makes it the first actor in the history of commerce with no inherent identity. This is not a further detachment of identity from an individual but rather a complete severance. AI agents are not crude bots executing scripts. They negotiate, persuade, make complex decisions, and interact in ways indistinguishable from humans. Sophistication creates an illusion of legitimacy that masks the absence of any accountability behind the interaction.

The combination threatens to slow institutional adoption of agentic commerce for anything other than simple payments: blockchain gave us trusted transactions without knowing who the parties are. Smart contracts gave us execution without the legal infrastructure of contract. AI agents give us sophisticated autonomous actors with no identity, no accountability, and no connection to the legal systems that make commerce work.

Trust in depth is the framework for creating those foundational connections.

## III. Trust in Depth

In information security, "defense in depth" is the principle that layered security mechanisms compose into strong security even though each individual layer is imperfect. No single firewall, perimeter, or control is the sole protection. The strength is in the composition.

We propose a parallel concept for the agentic era: **trust in depth.**

Trust in depth is the principle that layered trust mechanisms—each imperfect on its own—compose into robust trust when applied together. No identity verification, attestation, or legal framework needs to be independently sufficient. In combination, the layers create a trust architecture that is resilient, proportional, and resistant to the attacks that break single-layer systems.

The framework requires four layers.

### Layer 1: Human Identity

At the base of every chain of agency, a human must be identifiable. The mechanism for establishing human identity can vary by context—government credential, biometric, KYC through a financial institution. But the layer is non-negotiable. It is the anchor for everything else.

The most robust human identity infrastructure today is payments. Except for pure cryptocurrency transactions, all payments trace back to financial institutions that maintain the strongest identity systems in existence; banks, credit bureaus, government agencies, KYC and AML compliance, are all interconnected and subject to extensive regulation. A single mobile payment involves five layers of identity verification in a sub-second interaction: the issuing bank's KYC, the card network's enrollment, tokenization verification, device biometric authentication, and hardware-secured credential storage.

But the identity is owned by the intermediary, not the individual, and the payment rail an agent uses determines its governance posture entirely. Implement Stripe, and the agent inherits KYC, accountability, and legal enforceability. Pay with a pure cryptocurrency wallet, and the identity chain is severed; there is no person, no entity, no jurisdiction, and no recourse when something goes wrong. The same economic action is governed radically differently depending on which payment rail is wagging the governance dog.

## Layer 2: Entity Attestation

The organizational structures on behalf of which a human operates must be verifiable and attributable. The entity must be connected to identifiable humans and a recognized legal structure, whether through corporate filings, DNS-based identity by implication, or sovereign digital credentials.

Entity identity reduces to human attestation. At the end of every chain of corporate filings, board resolutions, and certificates of good standing, a human signs something and asserts: "I am authorized to act for this entity." The system trusts that assertion, verified against a government record and an identity check. "Identity by implication" offers a practical shortcut; it associates a public key with a DNS record to create a lightweight, verifiable chain of inference. It is weaker than cryptographic attestation, but far more practical for routine interactions.

The space between personal identity and organizational authority is where the trust chain fractures most often. Business email compromise, built almost entirely on this gap, accounts for over $2.9 billion in reported losses annually in the United States alone (FBI IC3, 2023). For agents, the entity question is harder still: who is "the entity" behind an agent? The company that built it? Deployed it? The user who configured it? These questions have no precedent to make answers easy.

## Layer 3: Agreement Integrity

The agreement must be tied to a governing legal framework. Jurisdiction must be established and the terms to which the parties agreed must be recorded permanently, independently verifiable, and controlled by neither party.

Blockchain makes its highest-value contribution here. No other technology can permanently record the legal nexus of an agreement—parties, terms, jurisdiction, moment of consent—in a way that is independently verifiable, tamper-proof, and controlled by neither party. Agreement integrity is where all other layers converge. An agreement binds individual identity (who), entity identity (on whose behalf), terms, and jurisdiction into a single enforceable commitment. Without any one of these, the agreement is legally compromised. Blockchain cannot tell you who the parties are. But once identity is established through other layers, it can irrevocably record that those identified parties entered into these specific terms governed by this specific jurisdiction.

But in the current agentic landscape, agreements are reduced to transactions. Two agents complete an exchange, and whatever happened is considered agreed by virtue of the exchange, with no further information about the governing terms. This causes three fundamental problems. The first is temporal: real agreements carry obligations that extend beyond the moment of exchange like liability allocation, performance standards, breach remedies, implied warranties, method of recourse. None of this is captured when the agreement is the transaction and the transaction is the agreement.

The second problem is evidentiary. Even where agents negotiate terms beyond the immediate exchange, there is no reliable mechanism for recording those terms in a way that is binding on both parties. Does one agent record the agreement? Where? Does each agent maintain its own record? If so, which is authoritative when they disagree? The absence of a neutral, independently verifiable record means that any terms an agent negotiates are only as durable as the infrastructure of the counterparty's agent, which is to say, not durable at all.

The third problem is intent. Contract doctrine assumes that a human deliberated: weighed the terms, assessed the counterparty, and decided to commit. Agents operating at scale and speed currently sever that analog process. An agent may apply policy parameters set by a human, but how can that be known and trusted in retrospect? The agent's autonomy is what makes it useful and it is also what currently breaks the assumption of human intent on which agreement enforceability rests.

## Layer 4: Agent Authorization

The autonomous system acting on behalf of human principals must carry verifiable, bounded authority. The delegation chain, from organization to human to agent, must be auditable, revocable, and scoped.

Agent identity is new. The first three trust layers have centuries of precedent, decades of corporate registration history, and are backed by mature legal frameworks. Agent identity has no precedent, regulation, standards, or legal framework. The first authoritative government engagement, NIST's AI Agent Standards Initiative, including its Agent Identity and Authorization concept paper, arrived only in February 2026 (NIST, 2026). Yet infrastructure is being built at speed. Every major cloud platform has shipped dedicated agent identity systems in the past eighteen months, and none shares a common identity standard. Visa introduced its Trusted Agent Protocol in October 2025 as part of its broader Intelligent Commerce initiative, with over one hundred partners engaged across the ecosystem; Mastercard introduced Agent Pay the same month (Visa, 2025; Mastercard, 2025). The traditional financial system is extending its existing identity apparatus to agents without waiting for the technology industry to solve this trust problem.

But the moment a user delegates signing authority to an agent, it becomes critical to confirm that delegation, scope it, and build infrastructure that can hold the user accountable. If an agent acts outside its scope, who is liable? The delegation chain is complex: an organization authorizes a human, who configures an agent, which may in turn invoke sub-agents or third-party services. Each step attenuates the connection to the original authority. The question of scope is not only whether the agent was authorized, but whether each link in the chain was authorized, and whether the cumulative effect of delegated decisions still reflects the principal's intent. Adding to the uncertainty, government digital identity programs across the EU, India, South Korea, China, Japan, and the United States each reflect different values and political realities. This fragmentation

reflects genuine differences in trust models and jurisdictional sovereignty rather than a transitional phase.

The question is not “who is this agent?” An agent has no independent legal existence. The question is “whose authority does it carry, and what are the bounds?”

That legal framing does not eliminate the practical need for agent-level differentiation. Within any organization of meaningful scale, agents will specialize—by role, domain, skill, and track record. Counterparties will need to distinguish between them in the same way patients distinguish among doctors at a hospital. Legally, a doctor practices under the hospital's license and the hospital carries ultimate accountability; operationally, patients choose specific doctors based on individual specialization, credentials, and reputation. A company running a thousand agents will deploy them to distinct roles, and the parties those agents interact with will need to identify, compare, and select among them on their individual merits, not only on the authority of the organization behind them.

Standards-track efforts have begun to build the infrastructure for this. ERC-8004 ("Trustless Agents"), drafted in August 2025 and deployed to Ethereum mainnet in January 2026, defines three on-chain registries—identity, reputation, and validation—extending the Agent-to-Agent protocol (De Rossi et al., 2025). Together they give an agent a portable identifier, a verifiable track record, and an independently checked capability profile. The "trustless" label describes the mechanism (the registries operate without a central authority) not a rejection of the broader trust chain. In a trust-in-depth architecture, agent-level registries like these compose with organizational delegation rather than replacing it: the delegation chain establishes whose authority the agent carries and what the bounds are; the agent registry establishes *which* specific agent, with what history and what verified capabilities, is acting within those bounds.

## What the Composition Achieves

No single trust layer can deliver what the agentic era requires. But the combination can.

**Anchoring to real-world identity.** A verifiable chain from agent action to human principal—not just a wallet address, but a person or entity that can be found, communicated with, and held accountable.

**Legal enforceability.** Agreements that are legally enforceable because identity, consent, and terms are all provable. A cryptographic signature is technically sound but legally meaningless without a framework that defines who may sign, what the signature represents, and what consequences follow.

**Jurisdictional linkage.** Every transaction must be connected to a governing legal framework. Without jurisdiction, there is no law to apply, no arbitral institution or court to decide a dispute, and no enforcement mechanism.

**Recourse.** When things go wrong (and they will) there is someone to find, a record to examine, and a legal system to invoke. Without recourse, agreements are only useful if there is no dispute.

### Economic Deterrence

Trust in depth restores the economic deterrent that single-layer systems have lost. A bad actor or a rogue agent would need to compromise all four layers simultaneously: forge a human identity, fabricate an entity, manipulate the agreement record, and circumvent the authorization chain. The cost of doing so recreates structural friction: not the cumbersome, slow friction of analog processes, but layered verification that makes gaming the system expensive again.

## IV. Implementation Principles

Trust in depth is a framework, not a product. Its value will depend on how it is implemented. Three principles should govern that implementation.

### Built With Institutions, Not Against Them

The agentic future will be built on the same foundation as everything else: social constructs, translated for a new era. The legal system, imperfect, evolving, jurisdiction-specific, is foundational to enabling billions of people to cooperate at scale. Code can execute logic. It cannot provide justice, interpret intent, balance competing interests, or adapt to circumstances its authors did not foresee.

The institutions that have spent decades building trust infrastructure — arbitral institutions, financial regulators, standards organizations — are critical partners to an agentic future. The framework must extend their authority into the agentic space. Arbitral institutions can extend existing procedural frameworks to govern agent-to-agent disputes, applying established rules of evidence and remedies to novel fact patterns without building new legal rules from scratch. Financial regulators can extend existing identity and compliance requirements to the agent authorization layer. Standards bodies can define the interoperability protocols that connect these layers. The infrastructure exists; the task is extension, not invention.

### Open, Not Owned

Solutions at every layer must be open rather than owned by a single platform or company. They must interoperate across jurisdictions, technologies, and agent platforms. And they must accommodate a permanently fragmented identity landscape.

The key insight is a separation of concerns: the protocol defines what questions must be answered about identity at each layer; the provider determines how. A government digital ID, an enterprise identity platform, a blockchain credential, and a payment rail's KYC can all fulfill the same identity operation through different mechanisms. The protocol needs to

understand what was proven and at what assurance level; it does not have to understand how.

This architecture allows each layer to accommodate multiple approaches; unification is not coming, and any framework that depends on it will fail.

### Proportional, Not Maximal

Trust in depth does not require maximum assurance at every layer for every interaction. The level of trust must be proportional to the stakes of the transaction.

The EU has already codified this principle for the identification layer. The eIDAS regulation defines three tiers of electronic signature: Simple (a typed name suffices), Advanced (uniquely linked to the signer, under their sole control), and Qualified (backed by a regulated trust service provider, carrying the legal weight of a handwritten signature). This legal framework is already in place and governing hundreds of millions of transactions. Trust in depth applies the same principle across the full identity stack.

Every approach to bridging physical and digital identity trades somewhere on the spectrum between convenience and integrity. Corporate-managed identity (processors, banks) provides strong identity at the cost of centralization. A cryptocurrency wallet provides speed and autonomy at the cost of accountability. Government-issued digital credentials are the closest thing to closing the gap. Reputation-based systems complement identity but do not substitute for it; they tell you an actor has behaved well in the past, but not who that actor is.

Trust in depth does not require one approach. It requires that whatever approach is chosen, all four layers are addressed. And the assurance level at each layer should match the stakes: a two-dollar API call and a five-million-euro contract do not need the same infrastructure.

## V. Conclusion

History shows that new technologies force the evolution of legal and social frameworks, not the other way around. The printing press led to copyright law. The automobile led to traffic regulation and liability insurance. The internet led to electronic commerce legislation and data privacy regulation. The law adapts with each technology innovation, but only after a period of disorder that could have been shortened by earlier engagement.

AI agents are the next forcing function. The disorder is already here and the frameworks are not.

Trust in depth — layered, composable, proportional trust mechanisms spanning human identity, entity attestation, agreement integrity, and agent authorization — is a starting point. It is a principled architecture for finding the best answers. The alternative is not a world without rules. It is a world where the rules are written by whoever moves fastest.

## References


De Rossi, M., Crapis, D., Ellis, J., & Reppel, E. (2025). "ERC-8004: Trustless Agents [Draft]." *Ethereum Improvement Proposals, no. 8004*. https://eips.ethereum.org/EIPS/eip-8004

Federal Bureau of Investigation. (2023). *2023 Internet Crime Report*. Internet Crime Complaint Center (IC3). https://www.ic3.gov/AnnualReport/Reports/2023_IC3Report.pdf

Imperva. (2025). *2025 Bad Bot Report*. Thales Group. https://www.imperva.com/resources/resource-library/reports/2025-bad-bot-report/

Mastercard. (2025). “Agentic Token Framework: Driving Trusted AI Transactions.” https://www.mastercard.com/global/en/news-and-trends/stories/2025/agentic-commerce-framework.html

National Institute of Standards and Technology. (2026). “Announcing the AI Agent Standards Initiative for Interoperable and Secure Innovation.” https://www.nist.gov/news-events/news/2026/02/announcing-ai-agent-standards-initiative-interoperable-and-secure

Plesner, A., Vontobel, T., & Wattenhofer, R. (2024). “Breaking reCAPTCHAv2.” *arXiv:2409.08831*. https://arxiv.org/abs/2409.08831

Szabo, N. (1996). "Smart Contracts: Building Blocks for Digital Markets." EXTROPY: The Journal of Transhumanist Thought, no. 16. https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart_contracts_2.html

Visa. (2025). “Visa and Partners Complete Secure AI Transactions, Setting the Stage for Mainstream Adoption in 2026.” https://usa.visa.com/about-visa/newsroom/press-releases.releaseId.21961.html