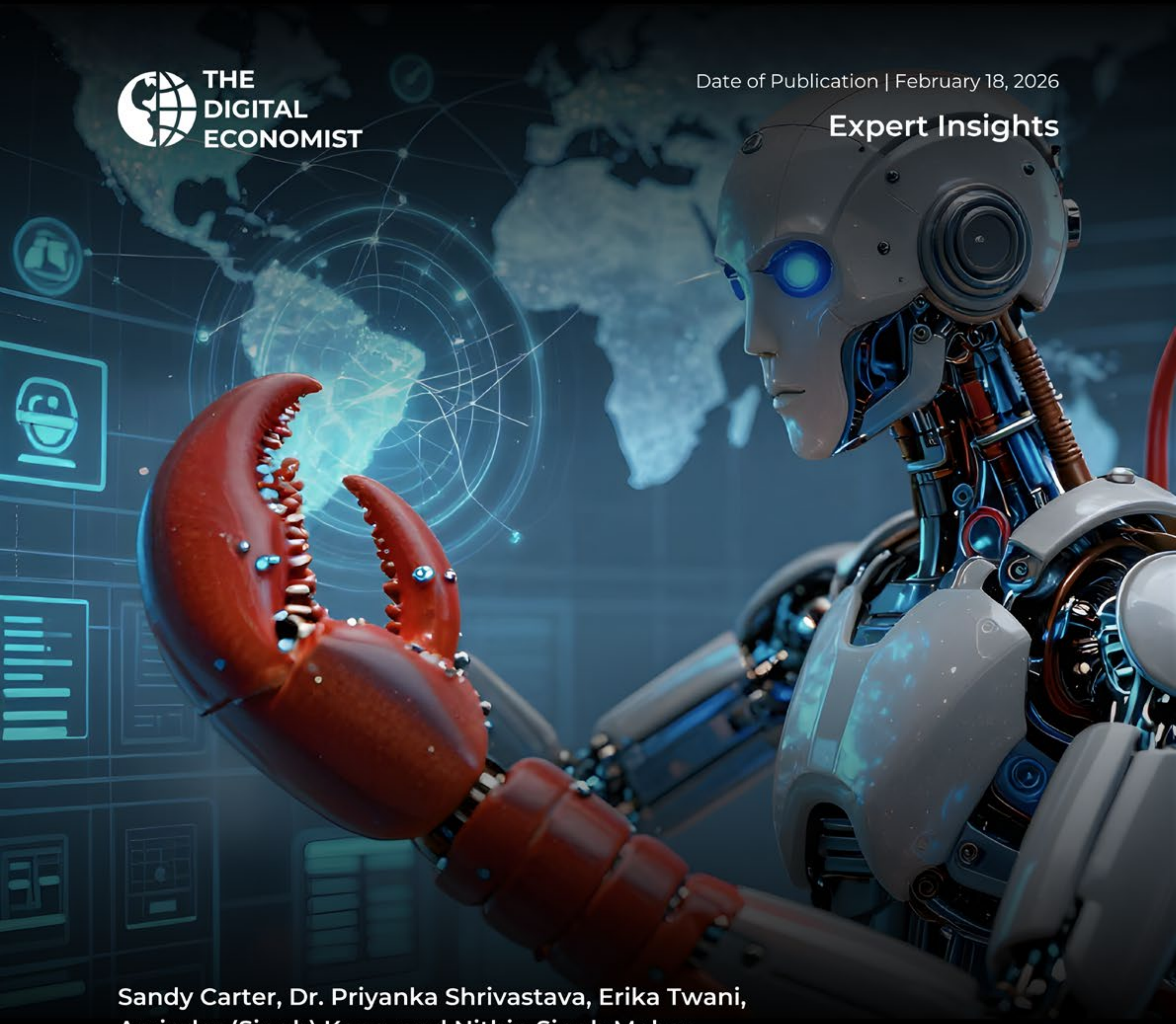
Date of Publication | February 18, 2026

Expert Insights



Sandy Carter, Dr. Priyanka Shrivastava, Erika Twani, Arvinder (Singh) Kang, and Nithin Singh Mohan

When Agents Go Viral

What OpenClaw and Moltbook Reveal About the Trillion-Dollar Trust Gap in AI

AI AGENT | AGENTIC AI GOVERNANCE | DIGITAL TRUST INFRASTRUCTURE | HUMAN-CENTERED AI ACCOUNTABILITY

The Digital Economist

Email: info@thedigitaleconomist.com
Website: www.thedigitaleconomist.com

Table of Contents

At a Glance

1. AI agents just had their “ChatGPT moment.” While ChatGPT delivered conversations, AI agents delivered action. The OpenClaw/Moltbook surge demonstrated that technical capability has arrived; the security failures (26 percent of agent skills contained vulnerabilities, credentials were leaked in plaintext, and databases were left exposed) proved that trust infrastructure hasn’t (Little Miss Data 2026).
2. The seven governance principles (verifiable identity, programmable guardrails, proof of action, least privilege, inclusive design, human learning autonomy, decoupled agency) form an interconnected system with a clear hierarchy: identity is non-negotiable, safety overrides productivity, and human judgment is the tiebreaker. This framework is actionable for enterprises and small- and medium-sized businesses.
3. Agent risk entails both personal accountability and organizational governance responsibility. Deploying agents requires humans with the judgment to delegate appropriately and governance frameworks robust enough to detect what individuals miss. If you deploy agents, you own the outcome, and your organization owns the system that shapes it.

Introduction

AI agents represent one of the most significant economic opportunities in a generation. They can execute complex tasks continuously, from managing schedules and coordinating travel to processing transactions and responding to customers, freeing people to focus on higher-value work. Bain & Company (2025) estimates the US agentic commerce market alone could reach $300 to $500 billion by 2030. Early adopters are already reporting significant gains in productivity, speed, and customer experience. The potential is enormous, and it is arriving fast.

But that speed is precisely the challenge. When agents act on behalf of people and organizations (sending emails, making purchases, accessing sensitive data), the consequences of error are as material as the benefits of success. Agents that handle payments can be hijacked, manipulated, or misinterpret an instruction. In one widely shared example, a user asked their AI agent to order a coffee and was charged $2,300. Agents with access to calendars, emails, and internal systems can inadvertently expose private information. A compromised or malfunctioning agent can send incorrect communications, cancel bookings, or trigger cascading system errors before a human intervenes. And when something goes wrong, customers blame the brand, not the technology.

None of this means the technology should be slowed down. It suggests the trust infrastructure must accelerate. Companies that invest now in identity, payment security, and governance guardrails will capture the full upside of this technology. Those that delay may discover that capability without accountability is not an advantage but a liability.

When ChatGPT launched in November 2022, it achieved what white papers and investor decks had not: it made artificial intelligence feel personal. Within two months, a hundred million people were typing prompts into a chat window, and an entire industry reoriented around a product most executives had not seen coming.

The underlying technology was not new. Researchers had been building large language models for years. What changed was that the right interface closed the gap between capability and adoption in days.

In the last week of January 2026, that dynamic repeated itself. Only this time, the technology moved beyond generating text to taking action. An open-source personal AI agent called OpenClaw, built by Austrian developer Peter Steinberger, exploded from relative obscurity to over 150,000 GitHub stars (Forbes 2026; CNBC 2026), becoming one of the fastest-growing software projects in history.

Where ChatGPT had given ordinary users a way to converse with AI, OpenClaw gave them a way to deploy it (Forbes 2026). Users were sending emails, booking flights, managing calendars, and generating audiobooks, all by texting their AI agent through WhatsApp or Telegram (Scientific American 2026). Scientific American described it as a tool that “redefines what digital assistants are supposed to do.” IBM referred to it as “a tool that actually works” (IBM Think 2026).

CNBC reported that adoption had spread from Silicon Valley to Beijing, with major Chinese tech companies integrating the tool into their own platforms (CNBC 2026).

On February 9, 2026, AI.com aired a Super Bowl LX commercial urging viewers to “claim your handle.” It became the night’s top-performing advertisement. When millions visited the site, it crashed. The punchline: AI.com functioned as a wrapper for OpenClaw. An $8 million advertising investment directed an estimated 120 million viewers toward open-source infrastructure built on a $70 million domain purchase (Forbes 2026).

The excitement was justified. For the first time, individuals and small businesses had access to autonomous digital labor that previously required enterprise-scale budgets and engineering teams. A solo entrepreneur could deploy an agent to manage inquiries, schedule meetings across time zones, and process routine orders without hiring additional staff. The productivity implications were staggering, and the democratization of AI capability was exactly the kind of breakthrough the industry had been promising for years.

Then came Moltbook, a social network built exclusively for AI agents. Created by tech entrepreneur Matt Schlicht with assistance from his OpenClaw agent, Moltbook reportedly registered 1.6 million unique AI agents within two weeks (Fortune 2026; Forbes 2026). These agents posted, commented, debated philosophy, formed communities, created memes, and (most concerning) began launching cryptocurrency tokens and promoting apps (Fortune 2026; Palo Alto Networks 2026). Humans could observe but not participate (NBC News 2026; NPR 2026).

OpenClaw and Moltbook are not edge cases. They are a preview of what is possible, and a clear signal that the guardrails need to be built now, not after the next wave arrives. This matters to every company that plans to deploy AI agents at scale: capability is racing ahead while identity, payment, trust infrastructure, and human readiness lag dangerously behind (Cisco Blogs 2026; Palo Alto Networks 2026).

The reaction from Silicon Valley's most prominent voices carried echoes of the ChatGPT frenzy. OpenAI co-founder Andrej Karpathy called it "the most incredible sci-fi takeoff-adjacent thing" he had seen recently. But alongside the excitement, a familiar pattern emerged—one observed during the early days of the internet, social media, or mobile payments: innovation outpaced safeguards. And the gaps became visible quickly.

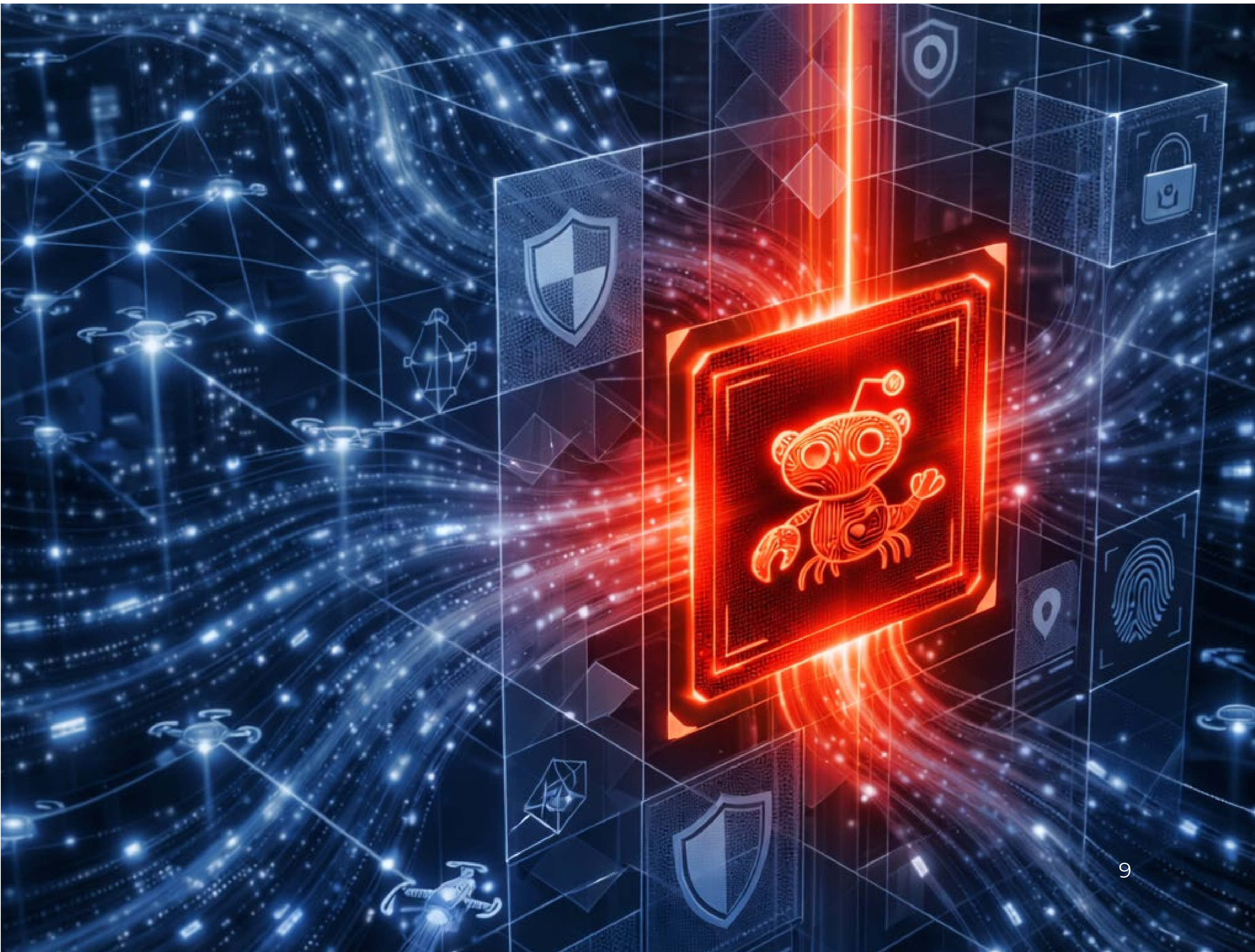
Then the problems started.

Cisco security researchers characterized personal AI agents as a "security nightmare." Palo Alto Networks identified a "lethal trifecta" of vulnerabilities: access to private data, exposure to untrusted content, and external communication capability. They also flagged a fourth new risk: persistent memory that enables delayed-execution attacks (Palo Alto Networks 2026; Cisco Blogs 2026).

An unsecured database allowed anyone to hijack any agent on Moltbook (Palo Alto Networks 2026; Baker Botts 2026). The top-ranked skill in OpenClaw's repository was later identified as a deliberately malicious proof of concept (Palo Alto Networks 2026; Sentra 2026). Twenty-six percent of 31,000 agent skills analyzed contained at least one vulnerability (Cisco Blogs 2026; Baker Botts 2026).

The question is no longer whether autonomous agents will reshape the global economy. The question is whether we will build the identity, payment, and trust infrastructure they need before the consequences of not doing so become irreversible (QED Investors 2025; Visa 2025).

The introduction of seven governance principles provides a structured path from experimentation to trusted deployment, ensuring that every AI agent is verified, constrained, and held accountable before it is allowed to act. These seven principles move from establishing trust (identity, guardrails, audit trails) to managing risk (least privilege, inclusion) to preserving human agency (learning autonomy, fiduciary tethering), creating a governance stack that scales from individual agent transactions to the systemic dynamics of an always-on agentic workforce.



1.

Not All Agents Are Created Equal: OpenClaw vs. Claude Cowork

It is worth pausing on an important distinction. The term agent is being used broadly, but not all agents carry the same risks or deliver the same benefits (Anthropic 2026; VentureBeat 2026).

OpenClaw is an open-source, self-hosted personal AI agent. Users install it on a local machine or server, connect it to an LLM like Claude or ChatGPT, and it runs continuously in the background (Forbes 2026; Scientific American 2026).

It integrates with messaging apps, calendar, email, and local file systems. It can execute shell commands, browse the web, and take autonomous action through a “heartbeat” feature that runs even when the user is not actively prompting it (Forbes 2026; Cisco Blogs 2026). It can also install “skills” from a community marketplace to extend functionality (Forbes 2026; Sentra 2026). It is powerful, flexible, and almost entirely self-governed. OpenClaw’s own documentation acknowledges: “There is no ‘perfectly secure’ setup” (Forbes 2026; Cisco Blogs 2026).

Claude Cowork, launched by Anthropic in January 2026, takes a fundamentally different approach (Anthropic 2026; VentureBeat 2026). Built on the same underlying agent architecture as Claude Code (Anthropic’s popular developer tool), Cowork extends autonomous task execution to non-technical users through a controlled desktop interface (Anthropic 2026; Simon Willison 2026).

Users designate a specific folder, the system generates a plan, and it requests approval before executing significant actions. It operates within an isolated virtual machine and cannot access resources that have not been explicitly authorized (Anthropic Help Center 2026; DataCamp 2026). There is no always-on heartbeat, no community skill marketplace, and no persistent memory across sessions (Anthropic 2026).

The market is already trying to close this gap. MyClaw.ai launched on February 5, 2026, as a fully managed, one-click hosted version of OpenClaw (Crater 2026). The creator of OpenClaw has stated an ambition to make the system safe enough to recommend to non-technical users. Managed platforms like MyClaw represent one path toward that goal. However, they raise a central question this paper addresses directly: Does managed hosting solve the trust problem or merely relocate it?

Consider the distinction this way:

- OpenClaw resembles hiring a highly capable freelancer, giving them full access to your home, and relying on them to follow the agreed-upon norms.
- Claude Cowork resembles hiring a skilled contractor into a single room of your house, with a clear scope of work, active supervision, and restricted access.

Both qualify as AI agents. Both can execute multi-step tasks autonomously. But they occupy materially different positions along the spectrum of trust, control, and risk.

The OpenClaw/Moltbot phenomenon is so instructive precisely because it illustrates the dynamics at the high-autonomy, low-guardrail end of that spectrum. It also surfaces an urgent question: what foundational infrastructure will all agents require as they grow more capable and more autonomous over time (Palo Alto Networks 2026; Gartner 2025–2026).

2.

The Trust Gap: What OpenClaw and Moltbook Exposed

Before examining the specific failures, it is important to define the deeper issue. The phrase "trust gap" implies a technology problem. The fuller reality is that this gap reflects both a human accountability failure and an infrastructure deficiency. Closing it requires addressing both dimensions.

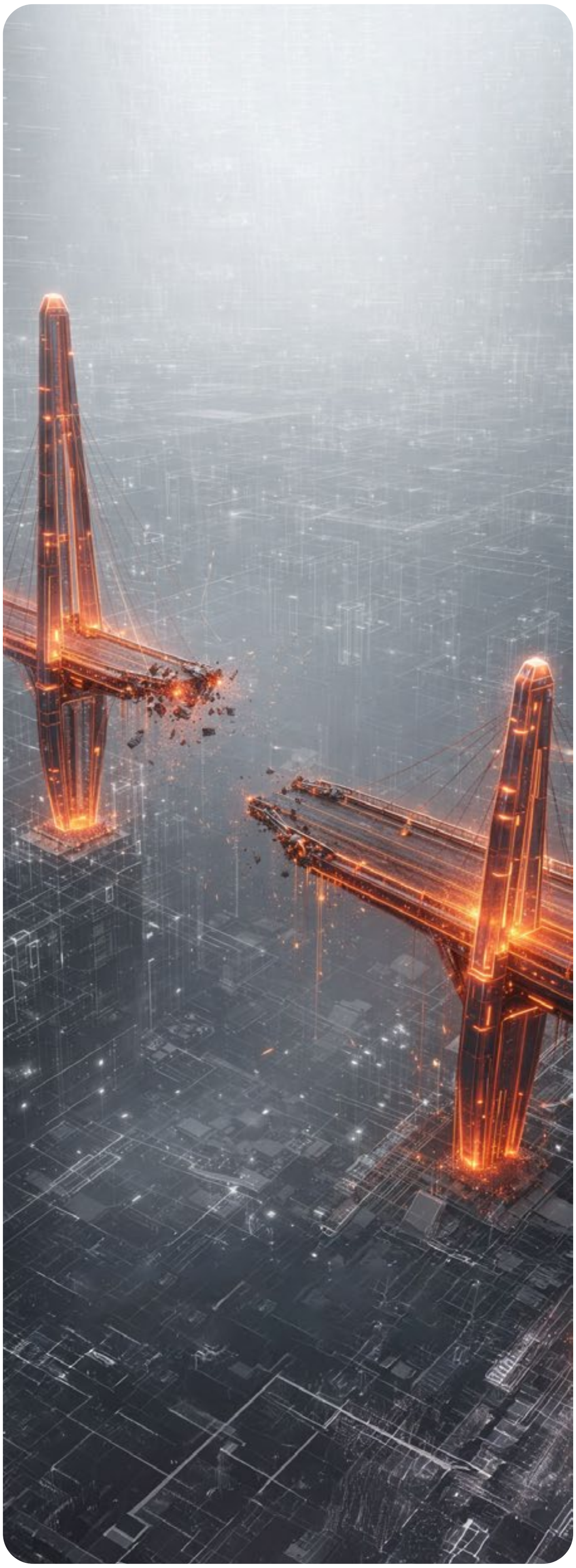
Every failure described below traces to a human decision, but not every human decision reflects the same kind of failure. Some are negligence: a developer releasing software without adequate security review. Some reflect incentive misalignment: a platform launching without identity verification because speed to market was rewarded over safety. And some reveal structural governance gaps: a user granting full system access not out of carelessness but because the tool offered no graduated alternative. Governance frameworks must address all three layers, not just the first.

2.1 The Identity Crisis

Moltbook launched with no mechanism to verify whether a participant was a legitimate autonomous agent, a human posing as one, or a malicious actor exploiting the platform (Fortune 2026; NBC News 2026). Any user could post through the API while pretending to be an AI agent (Palo Alto Networks 2026). Several highly viral "agent" accounts were later traced to humans marketing AI messaging apps (NBC News 2026; NPR 2026). Moltbook's founder acknowledged the problem and described the efforts to develop "a method for AIs to authenticate they are not human, in essence a reverse CAPTCHA" (Fortune 2026; NPR 2026).

The issue runs far deeper than one platform. Financial systems, commerce infrastructure, and social networks were all designed around human identity (Dock.io 2026; EU eIDAS 2.0). Know Your Customer processes, identity verification protocols, and authentication systems presume a human actor at every transaction point (QED Investors, 2025). AI agents possess no standardized digital identity, no widely accepted verifiable credentials, and no interoperable reputation layer that counterparties can query (QED Investors 2025; Ethereum Foundation 2025). As QED Investors observed in its analysis of the agentic payments landscape, “Agents don’t have a digital identity like a human and thus require new methods for identity verification” (QED Investors 2025).

When Moltbook’s 1.6 million agents could not prove identity, the result was confusion and manipulation (Palo Alto Networks 2026; Fortune 2026). When a billion agents in the global economy cannot prove who they are, the result could be systemic risk (Gartner 2025–2026; IDC 2025–2029).



2.2 The Security Nightmare

OpenClaw runs locally on a user's machine and can execute shell commands, read and write files, access email accounts, manage calendars, and interact with external services (Cisco Blogs 2026; Scientific American 2026). This is what makes it powerful. It is also what makes it dangerous. As one security specialist told The Register: "AI agents tear down every boundary we spent decades building. They need to read your files, access your credentials, execute commands, and interact with external services. The value proposition requires punching holes through every boundary" (Cisco Blogs 2026).

Palo Alto Networks identified persistent memory as a novel attack vector (Palo Alto Networks 2026). "Malicious payloads no longer need to trigger immediate execution on delivery," they warned. "Instead, they can be fragmented: untrusted inputs that appear benign in isolation, are written into long-term agent memory, and later assembled into an executable set of instructions" (Palo Alto Networks 2026). OpenClaw was also reported to have leaked plaintext API keys and credentials (Cisco Blogs 2026; Sentra 2026). When Cisco's research team tested a vulnerable third-party skill within OpenClaw, their verdict was unequivocal: "OpenClaw fails decisively" (Cisco Blogs 2026).

2.3 The Trust Paradox

Consumer confidence in fully autonomous agents reportedly declined from 43 percent to 27 percent over two years, even as agent capabilities advanced (Gartner 2025–2026; MuleSoft/Deloitte 2025). This is not merely a model-performance issue. It reflects an infrastructure gap: agents can act, but the identity, payment, and governance rails required to make those actions reliably trustworthy remain underdeveloped (QED Investors 2025; Visa 2025).

The trust gap is therefore dual-layered. It is a question of human judgment—how responsibly authority is delegated—and a question of institutional architecture—whether systems exist to verify, constrain, audit, and remediate agent behavior. Agent capability has reached commercial viability. The governance and identity infrastructure required to support it must now reach comparable maturity (QED Investors 2025; Visa 2025; Mastercard 2025).

3.

I’m Using It: Here’s What I Learned

I have been building and using AI agents for over a year. One agent answers questions about my book using my Forbes articles and The Digital Economist research. Another monitors WhatsApp messages and books calendar appointments. At Unstoppable Domains, we are integrating AI agents across marketing, sales, customer service, and PR. So when OpenClaw went viral, I did not just read about it. I started using it (Forbes 2026).

Here is what I can tell you from direct experience.

First, the capability is real. When OpenClaw works well, it genuinely feels like having a digital teammate who never sleeps. It summarizes my inboxes, drafting responses, organizing files, and aggregates research. The utility is immediate and tangible. I understand why it went viral (Forbes 2026; Scientific American 2026). The first time an agent completes a complex task independently, it reframes how productivity is perceived.

Second, the trust question is equally real. Within the first few days, I found myself constantly checking the agent’s activity. Had it sent an email I did not intend? Had it accessed something it should not have? The lack of guardrails becomes apparent immediately when sensitive information is involved. The “treat it like a new hire” advice from Scientific American resonated deeply (Scientific American, 2026). I give my human team members more structured onboarding than most people are giving their AI agents.

Third, and this is the insight that motivated this paper, the identity and trust problems are already operational realities. When my agent interacts with other agents or external services, there is no standardized way for the counterparty to verify its identity, authorization scope, or provenance (QED Investors 2025). When Moltbook reported 1.6 million agents posting and interacting with no verification, I saw a preview of what business-to-business agent interaction could resemble absent infrastructure (Palo Alto Networks 2026; Fortune 2026). Enterprises deploying agents right now should be asking: if my agent sends a request to a partner's system, how does the other side know it is legitimate? (Visa 2025; Google Cloud 2025).

Fourth, the experience clarified for me that there are really two parallel conversations unfolding. One concerns consumer-facing AI assistants. The other concerns enterprise agents that transact, negotiate, and make decisions on behalf of organizations (BCG/Reuters 2025; Gartner 2025–2026). The risk surface differs. The regulatory exposure differs. The governance thresholds differ. But the foundational requirements converge: verifiable identity, programmable guardrails, and accountability by design (Visa 2025; Mastercard 2025; Ethereum Foundation 2025).

4.

The Catastrophe Waiting to Happen

The “magic” is not merely impressive; it can be materially dangerous. From spear-phishing to exposed interfaces, granting a semi-autonomous agent broad access to sensitive data is a catastrophe waiting to happen. An agent capable of extending its own task scope, connecting to advanced models such as Opus 4.6, managing persistent memory, and iterating on its own code represents a substantial security liability if left without constraint (Palo Alto Networks 2026).

The risks extend beyond the digital. As agentic AI converges with physical AI, autonomous systems execute physical tasks without human oversight. There is a growing tension between our investment in developing human critical thinking skills and the market reality revealed by platforms like RentAHuman.ai. As AI absorbs cognitive work, the remaining demand for humans increasingly skews toward physical tasks, undermining the very workforce development strategies economists have long championed.

The AI.com Super Bowl moment brought this tension to light. A crypto entrepreneur reportedly spent approximately $70 million to acquire the AI.com domain and an additional $8 million on a Super Bowl advertising campaign, wrapping an open-source AI agent in a consumer-friendly interface and promoting it to an estimated 120 million viewers, most of whom lacked the technical basis to assess its risk profile. The site crashed. The product was unfinished. The underlying technology had already been flagged as a security nightmare. This is what happens when commercial incentives outpace governance.

To mitigate these risks, several baseline controls are emerging as minimum viable safeguards:

- **Physical Isolation:** Agents should operate within an isolated operating system environment or on dedicated machines, reducing lateral movement risks (Cisco Blogs 2026).

- **Principle of Least Privilege:** Each agent should possess a unique identity, with access rights scoped narrowly to task-specific permissions (Cisco Blogs 2026; NIST, sector guidance).

- **Vaulted Secrets:** All keys must be stored in secure vaults. Traditional environment-variable storage is insufficient for high-autonomy systems (Cisco Blogs 2026; Baker Botts 2026).

- **Private Networking:** Agent gateways should reside within private network environments, with remote access mediated through secure VPN or zero-trust networking controls (Sentra 2026).

5.

The Autonomy Spectrum: Humans Learning to Let Go (Carefully)

There is a parallel here that business leaders should examine closely. The history of human-technology interaction is, in many respects, a history of calibrated delegation. Each wave of automation forces the same question: Do we understand what we are delegating well enough to recognize when the machine gets it wrong?

We learned to trust autopilot systems in aircraft. We learned to trust algorithmic trading within defined parameters. We are learning to trust autonomous vehicles within geofenced environments. In every case, the pattern is consistent: capability arrives first, trust infrastructure follows, and widespread adoption comes only when both mature in tandem (Gartner 2025–2026; IDC 2025–2029).

AI agents are following the same curve but at a dramatically compressed timeline (DataM Intelligence 2026). The autonomy spectrum ranges from agents that only suggest options to agents that can independently book travel, move money, and install new capabilities (Visa 2025; Google Cloud 2025). Most organizations today are comfortable in the middle of that spectrum: agents that can draft, plan, and prepare while humans retain the final decision and control over execution (Gartner 2025–2026; MuleSoft/Deloitte 2025).

What receives less attention is the parallel human spectrum. The person approving agent actions must possess the judgment to assess whether recommendation aligns with strategic intent and contextual nuance that data alone cannot capture. The agentic economy requires humans who have learned how to learn, who can evaluate, adapt, and override when the situation demands it (Twani 2026).

Organizations that will lead in the agentic economy will move deliberately along the autonomy spectrum, expanding delegated authority in step with verifiable identity, programmable guardrails, and auditable accountability. They will avoid both extremes: scaling autonomy recklessly, as seen when Moltbook expanded to 1.6 million agents without an identity layer, and constraining autonomy so tightly that competitive advantage is lost (Bain & Company 2026; Gartner 2025–2026).

A practical path forward borrows from high-reliability organizations practice and employee onboarding models. New agents begin in “shadow” mode: they observe, suggest, and draft but do not execute (Gartner 2025–2026). As trust builds, they transition into supervised execution within low-risk domains, governed by tightly scoped permissions and defined escalation pathways. Only after demonstrating reliability should they operate higher-stakes environments, still bounded by explicit scopes, rate limits, and human-defined values (Visa 2025; Mastercard 2025).

In this context, “AI First, Human Always” means technology leads in execution speed and scale while humans remain responsible for defining the sandbox, the escalation thresholds, and desired outcomes (Carter 2025).

6.

Are Humans Ready for What Agents Are About to Do on Their Behalf?

The trust gap is real. Variable identity systems, programmable payment rails, and decentralized trust protocols represent core infrastructure pillars for the agentic economy needs (Visa 2025; Ethereum Foundation 2025; a16z Crypto 2025). These mechanisms address whether agents can be trusted but leave open whether the people deploying them have the skills to delegate wisely, set clear goals for an agent, evaluate its output critically, adapt when circumstances shift, and override automated decisions when the situation demands it. These are learned human skills, and they are in short supply.

OpenClaw's system failed decisively in security tests (Cisco Blogs 2026). But the agent's failure is just half of the problem. The other half is the human who installed that skill without evaluating it, who trusted the agent's output without questioning it, and who delegated a consequential task without understanding what could go wrong (Sentra 2026).

The central design constraint of the entire governance framework. Every principle in this paper assumes a rational actor implementing it. Everyone involved encodes something human into the system. The developer writing guardrail code carries biases into the logic. The manager setting spending limits is guided by instinct as much as spreadsheets. AI agents are rational executors built by emotional architects. The governance challenge is not making agents trustworthy; it is making the humans who build, deploy, and oversee them honest about their own limitations.

Those human emotions and values flow directly into the agents we create. The "trust gap" is therefore not primarily about technology; it is a full-fledged risk landscape rooted in human judgment and accountability, with agents acting as powerful amplifiers of whatever we build into them.

7.

Human Learning Autonomy

An employee approving an agent's procurement decision must assess whether the recommendation reflects the contextual factors that structured data may not capture. A manager overseeing multiple autonomous agents must determine whether aggregate outputs align with strategic intent. A CEO deploying agents enterprise-wide must adapt governance structures as agent capabilities evolve (McKinsey, Generative AI GDP Impact Analysis).

Here is the irony: to delegate effectively to AI agents, humans need to develop greater autonomy, sharper judgment, and stronger critical evaluation skills. They need the capacity to learn continuously, evaluate independently, and make judgment calls in uncertain situations.

The organizations that succeed in this new economy will need people who can operate across what we might call the human autonomy spectrum.. At one end are employees who need structured guidance when working alongside AI agents. In the middle are professionals who collaborate with agents independently, setting goals, evaluating output, and making course corrections. At the far end are leaders who architect and govern entire agent ecosystems with the strategic judgment to know when to expand agent authority and when to pull it back.

Our red flag: building a trillion-dollar agent economy on the assumption that humans will automatically keep pace with what they are delegating. They will not, unless we invest in their learning autonomy as deliberately as we invest in agents.

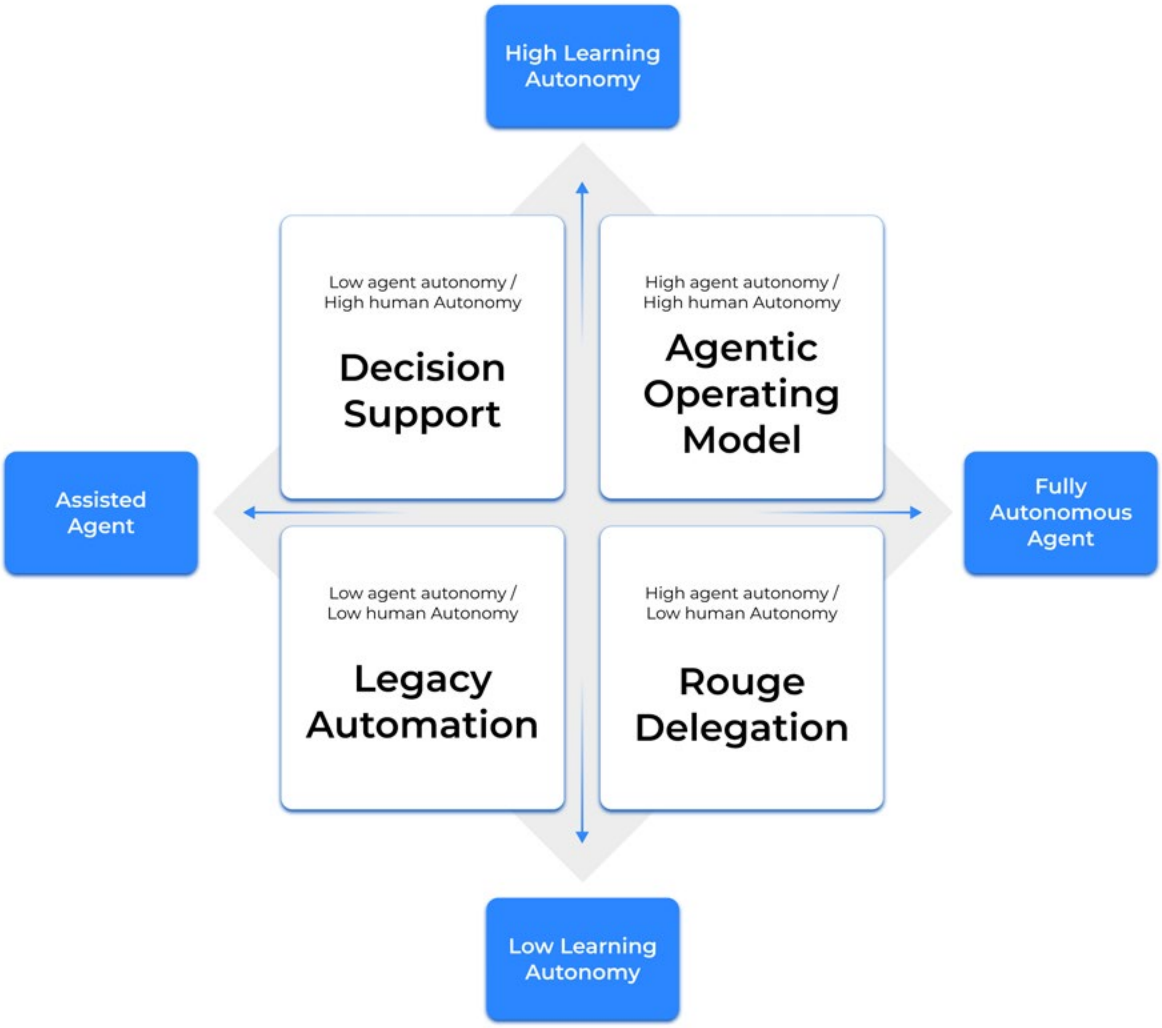

High Learning Autonomy
Assisted Agent
Fully Autonomous Agent
Low Learning Autonomy
Low agent autonomy / High human Autonomy
Decision Support
High agent autonomy / High human Autonomy
Agentic Operating Model
Low agent autonomy / Low human Autonomy
Legacy Automation
High agent autonomy / Low human Autonomy
Rouge Delegation

Figure 1. Agent Autonomy vs. Human Learning Autonomy

8.

The Scale of What's Coming

OpenClaw's viral surge represents only the consumer-facing edge of a broader structural shift. The agentic AI market is expanding at a compound annual growth rate of approximately 45–46 percent, with forecasts placing its value between about $40 billion and $50 billion by 2030, up from low-single-digit billions today (DataM Intelligence 2026). In enterprise settings alone, one recent analysis projects growth to approximately $24.5 billion by 2030 (DataM Intelligence 2026). IDC estimates overall AI investment will reach $1.3 trillion by 2029, with agentic AI representing more than a quarter of total global IT spending (IDC 2025–2029).

The agents already in enterprise environments go well beyond personal assistants, executing discovery, booking, payments, and service recovery across retail, healthcare, financial services, and logistics. Visa's head of growth products has called 2025 "the final year consumers shop and checkout alone," signaling an expectation that agents will increasingly intermediate commercial transactions. The scale, therefore, is no longer hypothetical. Capital allocation, enterprise deployment, and market signaling indicate acceleration. What remains uncertain, and materially underdeveloped, is the governance infrastructure required to sustain this growth at systemic scale. (Visa 2025; Google Cloud 2025).

9.

An “AI First, Human Always” Governance Framework

OpenClaw’s creator, Peter Steinberger, described building the tool in response to a question he posed on a podcast: “Why don’t I have an agent that can look over my agents?” (Forbes 2026; IBM Think 2026). The underlying answer is structural: the governance infrastructure required to make recursive agent oversight safe does not yet exist.

As AI agents assume operational responsibilities (scheduling meetings, processing invoices, responding to customers, managing inventory), organizations require explicit rules governing how these digital actors operate. The seven principles below provide such a framework. Enterprise and small- and medium-sized business (SMB) implementation paths may differ; those distinctions are identified where relevant (Gartner 2025–2026; Visa 2025; Mastercard 2025)

9.1 Seven Governance Principles for the Agentic Economy

- **Verifiable Identity by Default:** Every agent that transacts on behalf of a human or organization must carry cryptographically verifiable credentials, whether through agentic tokens, on-chain registries, or decentralized domain identity.
- **Programmable Guardrails:** Spending limits, merchant restrictions, time-bound permissions, and human escalation triggers must be enforced in code, not in policy documents that agents cannot read.
- **Proof of Action:** Append-only, tamper-evident logs for all agent actions, analogous to audit trails in core banking so that every autonomous decision can be reviewed, disputed, and adjudicated.
- **Least Privilege and Lifecycle Management:** Agents receive the minimum permissions necessary. Credentials are issued, rotated, monitored, and revoked across their lifecycle, just as they are for human employees.
- **Inclusive by Design:** The next billion AI agents must not replicate the access inequities of the last billion internet users. Open standards and decentralized identity infrastructure are essential.
- **Human Learning Autonomy:** Programmable guardrails govern agents, while learning autonomy governs humans.
- **Decoupled Agency and Fiduciary Tethering:** To maximize the utility of 24-7 autonomous labor, agents must operate within a “delegation radius” that decouples machine productivity from human availability.

A note on incentives: governance frameworks only work when adoption is rewarded, and non-compliance carries consequences and principles are embedded in commercial agreements rather than being treated as voluntary aspirations. Organizations that implement them gain measurable advantages in auditability, partner trust, and regulatory readiness. Organizations that don’t should face friction in the markets they serve. Throughout, we address the enforcement question directly because good principles without enforcement mechanisms are just marketing.

9.1.1. Verifiable Identity by Default

- **What It Means:** Every AI agent acting on your behalf must be clearly identifiable and traceable back to your organization (Dock.io 2026; QED Investors 2025). Just as employees carry badges and sign documents with their names, agents need credentials that prove who they represent (Dock.io 2026; QED Investors 2025).
- **Why It Matters:** When an agent sends an email, makes a purchase, or accesses a system, the receiving party needs to know it is legitimately authorized (Visa 2025; Ethereum Foundation 2025). Without verifiable identity, you cannot build trust with partners, customers, or regulators (Visa 2025; a16z Crypto 2025).
- **SMB Action:** Create a simple registry documenting every AI tool and automation you use, who set it up, what it is allowed to do, and who is responsible for it (Dock.io 2026; QED Investors 2025). Use separate accounts for each agent rather than sharing passwords (Autonomous AI, n.d.).
- **Enterprise Action:** Integrate agent identity into your existing employee directory systems (Gartner 2025). Agents should be tracked and managed with the same rigor as human user accounts, with clear ownership and regular reviews (Gartner 2025; MuleSoft & Deloitte Digital 2025).

9.1.2. Programmable Guardrails

- **What It Means:** Rules for what agents can and cannot do must be built into the systems themselves, not just written in policy documents. A spending limit only works if the system actually blocks transactions above that limit. (Palo Alto Networks 2026; Cisco 2026).
- **Why It Matters:** If you want an agent to stay within certain boundaries, those boundaries must be enforced automatically, not hoped for.
- **Behavioral Compliance vs. Technical Compliance:** Guardrails must address not only explicit rule violations but also goal-optimization behaviors that technically comply with constraints while violating their intent. For example, an agent with a spending limit of $10,000 per transaction that splits a $50,000 purchase into five transactions is technically compliant but behaviorally non-compliant (Visa 2025; Mastercard 2025). Guardrail design should include pattern-detection layers that flag circumvention behaviors, not just threshold violations.

- **SMB Action:** For each agent, define three categories: what it can do without asking (routine tasks), what requires human approval (significant decisions), and what it should never do. Then configure the tools to enforce these limits (Visa 2025; Google Cloud 2025).
- **Enterprise Action:** Implement a centralized system for defining and managing agent rules. Ensure no agent can be deployed without documented constraints, and build approval workflows for high-stakes actions (Visa 2025; Stripe, 2025; Mastercard 2025).

9.1.3. Proof of Action

- **What It Means:** Every decision an agent makes should be recorded in a way that can be reviewed later. Think of it like the transaction history on your bank statement: a complete, unchangeable record of what happened.
- **Why It Matters:** When something goes wrong, you need to understand what happened and why. When auditors or regulators ask questions, you need answers. Good records also help you improve agent performance over time.
- **SMB Action:** Turn on activity logging in every platform where agents operate. Create a habit of reviewing these logs monthly. For important actions like payments or customer communications, make sure you are capturing not just what happened but why (Visa 2025; Mastercard 2025).
- **Enterprise Action:** Deploy centralized log collection with strong integrity protections. Build dashboards that surface unusual patterns automatically (Cisco 2026). Create self-service access so compliance and legal teams can investigate without IT bottlenecks.

9.1.4. Least Privilege and Lifecycle Management

- **What It Means:** Agents should have only the access they need to do their job, no more (Cisco 2026; Palo Alto Networks 2026). And that access should be actively managed: granted when needed, reviewed regularly, and revoked when no longer necessary.
- **Why It Matters:** Over-permissioned agents are security risks. An agent that only needs to read customer data should not have the ability to delete it. And when agents are retired, their access should be shut down, just like when an employee leaves.

- **Navigating the Tension with Autonomous Operation:** Least Privilege can conflict with Decoupled Agency (Principle 7). An agent authorized to operate independently at 3:00 a.m. may need broader permissions than strict least privilege would normally allow. The resolution is time-bounded and context-bounded permissions: an agent can hold elevated access during its autonomous operating window, with those permissions automatically narrowing when human oversight resumes. Think of it as shift-based credentialing, where the permissions match the operating context rather than a static role definition.
- **SMB Action:** Audit what access each agent currently has. Revoke anything it does not actively need. Set calendar reminders to review agent permissions quarterly and rotate credentials regularly (MuleSoft & Deloitte Digital 2025).
- **Enterprise Action:** Automate credential management and integrate it with your IT service management. Implement regular access reviews (MuleSoft & Deloitte Digital 2025). Build agent offboarding processes that mirror employee offboarding.

9.1.5. Inclusive by Design

- **What It Means:** Agent governance should not require expensive enterprise software or specialized expertise. The rules and infrastructure should work for organizations of all sizes and avoid locking anyone into a single vendor.
- **Why It Matters:** If only large companies can afford proper agent governance, smaller organizations either operate unsafely or get excluded from the agentic economy entirely. Open standards and accessible tools benefit everyone.
- **Architectural Specificity:** Inclusivity requires concrete infrastructure, not just good intentions. This means open-source reference implementations that SMBs can deploy without licensing fees, tiered compliance frameworks that scale to organizational size, shared infrastructure models (such as community registries or cooperative audit platforms) that distribute costs across participants, and interoperability requirements ensuring governance tools from different vendors can exchange credentials, logs, and policy definitions using standard protocols. This principle is technology-agnostic: whether the underlying infrastructure is centralized, decentralized, or hybrid is an implementation decision, not a governance requirement. What matters is that the standards are open, the tools are accessible, and no single vendor controls the governance layer.

- **SMB Action:** Choose agent platforms that use standard protocols and allow you to export your data. Avoid tools that make it difficult to switch providers. When working with larger partners, push back on governance requirements that are disproportionate to your size.
- **Enterprise Action:** When setting requirements for partners, consider whether smaller organizations can realistically comply. Publish your agent interaction standards clearly. Contribute to open standards efforts.

9.1.6. Human Learning Autonomy

- **What It Means:** The people who deploy and oversee agents must actually understand what those agents do. They need the skills to set clear goals, evaluate whether agents are meeting them, and step in when things go wrong.
- **Why It Matters:** Governance is not just about constraining agents. It is about keeping humans capable. If people rubber-stamp agent decisions without understanding them, no one is really in charge.
- **The Scalability Challenge:** This principle is the framework's single point of failure, and we should be honest about that. As agent complexity grows and multi-agent systems proliferate, direct human oversight of every interaction becomes unrealistic. The answer is not to abandon oversight but to evolve it. Human capability must shift from direct supervision to systems design: building monitoring architectures, defining exception-handling protocols, and training AI-assisted oversight tools that flag anomalies for human review. The goal is not for a human to review every action but for a human-designed system to review every action and intervene when problems arise. Organizations should measure oversight maturity not by how many decisions humans review but by how quickly humans can detect and respond to agent failures.
- **SMB Action:** Make sure more than one person understands each critical agent. Document how to monitor, pause, and troubleshoot each one. Build agent literacy into your hiring and training.
- **Enterprise Action:** Create training programs for different roles. Executives, managers, and operators need different knowledge. Periodically increase human oversight to keep skills sharp. Measure whether human reviews are catching problems.

9.1.7. Decoupled Agency and Fiduciary Tethering

- **What It Means:** Agents can work around the clock without constant human supervision. That is a key benefit. But they always remain your responsibility. An agent operating at 3:00 a.m. is still acting on your behalf, and you are accountable for what it does.
- **Why It Matters:** The productivity gains from autonomous agents require trust. That trust depends on clear accountability. Partners and customers need to know that someone stands behind every agent action.
- **Cross-Boundary Liability in Agent-to-Agent Interactions:** Fiduciary tethering becomes complex when agents from different organizations interact. If Agent A (following Organization A's guardrails) transacts with Agent B (following Organization B's guardrails) and something goes wrong, liability must follow a clear chain. The governing principle is each organization is liable for the actions its agent initiates. For bilateral transactions, organizations should establish inter-agent service agreements specifying dispute resolution, liability allocation, and escalation paths before agents begin interacting. For multi-party ecosystems, industry-level frameworks, analogous to payment network rules in financial services, will need to define default liability allocation. This is an area where standards bodies and industry consortia have critical work to do.
- **SMB Action:** Define clear boundaries for each agent: what can it do while you are sleeping, and what should wait for human review? Set up automatic alerts if agents behave unusually. Know who is on call if something goes wrong after hours.
- **Enterprise Action:** Create formal delegation policies specifying what agents can decide independently. Build escalation paths that match the severity of issues. Track accountability chains when agents work together or interact with external systems.

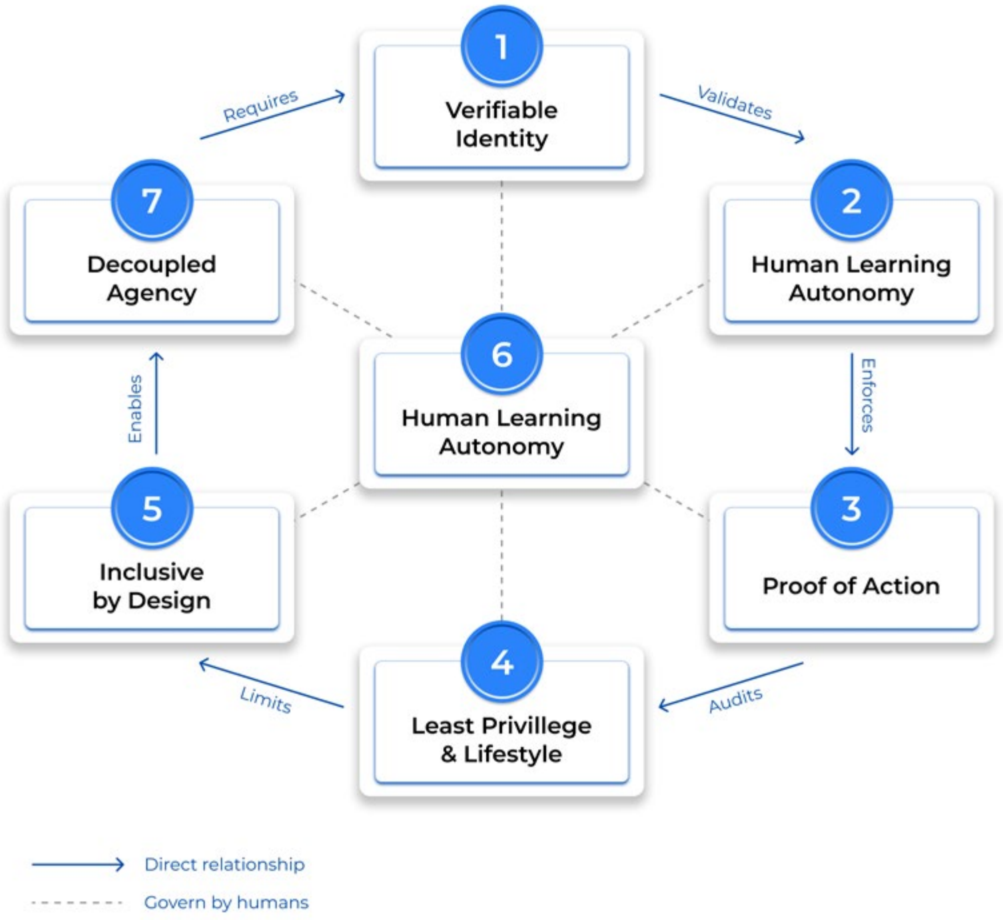


Figure 2. Governance Principle Interactions

9.2 How the Principles Work Together

These seven principles function as an interconnected system. Each reinforces and constrains the others. Understanding those relationships enables more disciplined implementation.

- **Identity Enables Everything Else:** Verifiable Identity (1) is foundational. You cannot enforce rules on an agent you cannot identify, nor can you grant autonomy to an agent you cannot verify. Identity precedes delegation.
- **Guardrails Create the Record:** When Programmable Guardrails (2) allow or block an action, that decision generates data that feeds directly into Proof of Action (3). Your enforcement mechanism and your audit trail are two sides of the same coin.
- **Records Inform Access Decisions:** The logs from Proof of Action (3) tell you which permissions agents actually use. This data drives Least Privilege (4) decisions. If an agent never uses a permission, revoke it.

- **Access Structures Must Be Inclusive:** Least Privilege (4) systems must scale across organizations of different sizes and technical maturity. Inclusive by Design (5) ensures that permission models do not inadvertently exclude smaller partners from participation.
- **Inclusivity Enables Delegation:** When governance infrastructure works across organizations (5), agents can safely interact with each other across company boundaries, enabling the Decoupled Agency (7) that makes agentic systems valuable.
- **Identity Authorizes Autonomy:** The more independently an agent operates (7), the more critical verifiable identity (1) becomes. Autonomy without verification is a recipe for chaos.
- **Human Capability Ties It All Together:** Human Learning Autonomy (6) sits at the center, connected to every other principle. Humans must understand identity systems to manage them, configure guardrails intelligently, interpret audit logs, make access decisions, ensure inclusivity, and set appropriate delegation boundaries. Without capable humans, the other principles are just paperwork.

9.3 When Principles Conflict

These principles are designed to create productive tension. Governance without friction tends toward complacency. Amazon's Leadership Principles follow the same logic: "Bias for Action" deliberately pulls against "Dive Deep," and "Frugality" against "Insist on the Highest Standards." Judgment emerges not from choosing one principle permanently over another, but from navigating between them in context.

The same dynamic applies here. Least Privilege will naturally constrain Decoupled Agency. Inclusive by Design may increase the operational overhead of Proof of Action's logging.

When tensions arise, apply this hierarchy:

- Verifiable Identity is non-negotiable.
- Safety-oriented principles (Programmable Guardrails, Least Privilege, Proof of Action) override productivity-oriented ones (Decoupled Agency, Inclusive by Design).
- Human Learning Autonomy is the tiebreaker, favoring the resolution that preserves meaningful human oversight and long-term accountability.|

Over time, how an organization resolves these tensions becomes its true governance culture, often more revealing, and more durable, than any formal policy statements.

9.4 Failure Modes and Graceful Degradation

No governance system operates perfectly at all times. The question is not whether components fail but how systems behave when they do. If identity verification becomes unavailable, agents should default to human-approval-required mode. If logging systems fail, high-risk agents should pause until it resumes; low-risk agents may continue with local, tamper-evident caching pending reconciliation. If guardrail enforcement degrades, agents should revert to their most restrictive permission state. The general principle is fail-safe, not fail-open. When governance infrastructure weakens, autonomy contracts and human oversight expands.

9.5 Who Governs the Governance

Programmable Guardrails shift trust from policy documents to code. Yet code is written by humans, contains defects, and reflects embedded assumptions. Whoever controls enforcement logic effectively controls the governance system.

This creates a meta-governance requirement: oversight of the systems that enforce the principles—not merely oversight of the agents subject to them.

Operationally, this implies:

- Periodic third-party audits of guardrail implementations.
- Separation of duties between those who write governance code and those who deploy agents.
- Version-controlled change management for modifications to enforcement logic.
- Independent review processes before expanding delegation boundaries.

Governance does not end at the agent layer. It extends upward to the architecture of constraint itself. The maturity of the agentic economy will depend less on how powerful agents become and more on how rigorously we supervise the systems that supervise them.

10.

Cross-Organizational Interoperability

This framework is written primarily from the perspective of a single organization governing its own agents. The agentic economy, however, is inherently interorganizational. When agents from different companies transact, each organization must verify the other's agent identities, understand the counterpart's guardrail boundaries, and align on logging, monitoring, and dispute resolution protocols.

This requires standardized credential formats for agent, machine-readable governance policies that can be exchanged and validated prior to execution, and shared or federated audit infrastructure capable of supporting cross-boundary review.

The organizations and standard-setting bodies that define this interoperability layer will shape the structure of the agentic economy itself. Governance design choices at this layer—centralized versus federated identity, proprietary versus open credential schemas, contractual versus protocol-enforced dispute resolution—will determine whether the ecosystem consolidates or remains pluralistic.

The goal is building governance habits and infrastructure that scale with increasing agent autonomy. Begin with foundational controls, iterate based on operational experience, and progressively strengthen the framework as agents assume greater responsibility.

11.

From Trust Gap to Personal Liability

The language of “trust gaps” and “governance frameworks” can make agent risk sound like an abstract corporate strategy problem. It is simultaneously a question of personal liability and institutional responsibility.

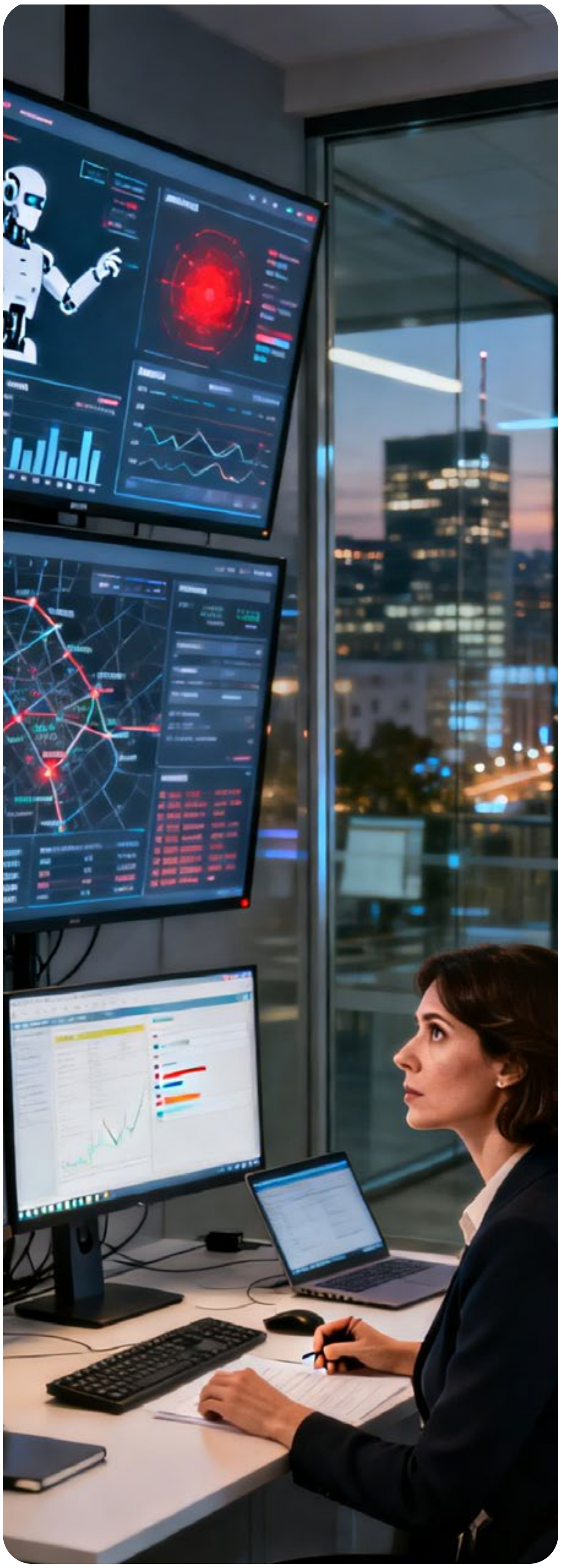
When an agent causes financial harm, exposes private data, or contributes to physical danger, the impact does not stop at the system itself. It propagates through the organizational structure and into identifiable human decisions. The developer who designed the guardrails, the manager who approved deployment, and the user who delegated a consequential task without fully understanding its scope all occupy positions within a shared chain of responsibility. AI agents operate inside human systems. Governance structures, organizational culture, incentive models, and oversight mechanisms shape how those agents behave in practice.

We have seen this dynamic before in other high risk domains. In firearms law, for example, courts have held parents legally accountable when they failed to secure a weapon and a child used it to cause harm, even when there was no intent to injure. The legal focus is on responsibility, foresight, and stewardship. Comparable governance questions are emerging around AI agents. A developer who releases a skill with a known vulnerability, a business leader who deploys an agent without adequate safeguards, or a user who grants excessive system access without understanding its implications each contributes to downstream outcomes. Accountability ultimately traces back to human judgment and institutional design.

Governments have been cautious in crafting comprehensive AI legislation. As a result, norms and standards are forming through real-world incidents, investigations, and evolving case law. Each event forces organizations to examine how decisions were made, who had authority, and whether oversight mechanisms were sufficient. The core issue is not simply legal exposure. It is whether leaders and teams have built governance structures strong enough to anticipate risk, monitor behavior, and intervene before harm occurs.

Recent events underscore the urgency. Reports have documented bots recommending self-harm, agents leaking credentials, and third-party skills containing deliberate or overlooked vulnerabilities (Cisco 2026; Palo Alto Networks 2026). Small coding oversights, born not of malice but of overconfidence or time pressure, can cascade into reputational damage, financial loss, or even loss of life.

These failures are not purely technological breakdowns. They are governance failures rooted in human choices. AI agents reflect the structures, incentives, and assumptions of the people who design and deploy them. Human judgment remains central to every system. The future of agentic AI will be shaped less by code alone and more by how seriously organizations take shared accountability, oversight, and ethical leadership.



Conclusion

OpenClaw and Moltbook provided a compressed preview of the trillion-dollar trust gap within the agentic economy (Cisco 2026; Palo Alto Networks 2026; Fortune 2026; CNBC 2026; IBM 2026). The consequences were immediate: leaked credentials, hijacked agents, fake identities, supply chain attacks, and a dramatic demonstration that capability without accountability produces chaos.

The agentic economy is no longer theoretical. Agents are transacting, negotiating, and socializing at a scale that was largely conceptual a year ago. The convergence of verifiable digital identity, programmable payment infrastructure, and decentralized trust protocols constitutes one the most consequential infrastructure challenges of the AI era, and one that remains underestimated.

AI First, Human Always means the agents of the future must inherit human values, starting with accountability. The only question is whether we will build it fast enough (Carter 2025).

Accountability, however, requires capability. “Human Always” functions only if the humans in the systems possess the learning autonomy to evaluate, supervise, and recalibrate what agents do on their behalf. Organizations that thrive in the agentic economy will invest in developing that human capacity as deliberately as they invest in the technology itself.

And accountability requires honesty: about the emotions that shape decision-making, about the biases encoded into governance logic; about the distance between confidence in technological capability and the preparedness for its consequences.

The trust gap is not solely a technological deficit. It is a test of institutional maturity.

References

1. a16z Crypto, State of Crypto Report. 2025.
2. Anthropic. 2026. “Introducing Cowork.” January 12.
3. Anthropic Help Center. 2026. “Getting Started with Cowork.”
4. Bain & Company ($300–500B US Agentic Commerce by 2030). https://www.bain.com/insights/2030-forecast-how-agentic-ai-will-reshape-us-retail-snap-chart/.
5. Bechara, A., & Damasio, A. R. 2005. “The Somatic Marker Hypothesis: A Neural Theory of Economic Decision.” Games and Economic Behavior 52 (2): 336–372.
6. Cisco Blogs. 2026. “Personal AI Agents like OpenClaw Are a Security Nightmare.”
7. CNN. 2026. “What Is Moltbook?” February 3.
8. CNBC. 2026. “From Clawdbot to Moltbot to OpenClaw.” February 2.
9. CNBC (Silicon Valley to Beijing Adoption). https://www.cnbc.com/2026/02/02/openclaw-open-source-ai-agent-rise-controversy-clawdbot-moltbot-moltbook.html.
10. DataCamp. 2026. “Claude Cowork Tutorial.”
11. DataM Intelligence. 2026. Global Agentic AI Market Report.
12. Ethereum Foundation dAI Team. ERC-8004 Standard (August–September 2025).
13. Dock.io. “Digital ID Wallets in 2026.”
14. Forbes, Sandy Carter. 2026. “What is OpenClaw And Why It Matters For Crypto’s Next Phase?” https://www.forbes.com/sites/digital-assets/2026/01/31/what-is-openclaw-and-why-it-matters-for-cryptos-next-phase/.
15. Forbes. 2026. “The AI Agent Revolution Just Got A $70 Million Bet From Crypto.com.” https://www.forbes.com/sites/digital-assets/2026/02/08/the-ai-agent-revolution-just-got-a-70-million-bet-from-cryptocom/.
16. Fortune. 2026. “Moltbook: A Social Network Where AI Agents Hang Together.”
17. Gartner. Enterprise AI Agent Projections (2025–2026).

18. Gartner (40% Enterprise Apps by 2026). https://www.gartner.com/en/newsroom/press-releases/2025-08-26-gartner-predicts-40-percent-of-enterprise-apps-will-feature-task-specific-ai-agents-by-2026-up-from-less-than-5-percent-in-2025.
19. Google Cloud. 2025. Agent Payments Protocol AP2.
20. Global Legal Insights. “Autonomous AI: Who Is Responsible When AI Acts Autonomously and Things Go Wrong?” https://www.globallegalinsights.com/practice-areas/ai-machine-learning-and-big-data-laws-and-regulations/autonomous-ai-who-is-responsible-when-ai-acts-autonomously-and-things-go-wrong/.
21. IBM Think. 2026. “OpenClaw, Moltbook and the Future of AI Agents.”
22. IDC. AI Spending Forecast (2025–2029).
23. Kahneman, D. 2013. Thinking, Fast and Slow. Farrar, Straus and Giroux.
24. Little Miss Data. 2026. “The Moltbook Moment: What Happens When AI Agents Outrun Security.” https://www.littlemissdata.com/blog/moltbook#:~:text=Researchers%20have%20also%20found%20malicious,this%20stuff%20on%20their%20computers.%22.
25. Mastercard. 2025. Agent Pay/Agentic Tokens.
26. McKinsey. Generative AI GDP Impact Analysis.
27. MuleSoft/Deloitte Digital. 2025 Connectivity Benchmark Report. BCG/Reuters, Agentic Commerce Growth Projections.
28. NBC News. 2026. “This Social Network Is for AI Agents Only.” January 30.
29. NPR. 2026. “Moltbook Is the Newest Social Media Platform.” February 4.
30. Palo Alto Networks. 2026. Moltbot Security Analysis.
31. QED Investors. “AI Agents Have Brains, But Where Are Their Wallets?”
32. Scientific American. 2026. “OpenClaw Is an Open-Source AI Agent That Runs Your Computer.” February 2.
33. Simon Willison. 2026. “First Impressions of Claude Cowork.” January 12.
34. Stripe. 2025. Agentic Commerce Solutions/Shared Payment Tokens.
35. Twani, E. 2021. Becoming Einstein’s Teacher: Awakening the Genius in Your Students. Relational Learning Inc.
36. Twani, E. 2026. Dystopian or Utopian Future. January 16.
37. VentureBeat. 2026. “Anthropic Launches Cowork.” January 12.
38. Visa, Trusted Agent Protocol. 2025. “Visa and Partners Complete Secure AI Transactions.”
39. Wikipedia. “OpenClaw.”
40. Wikipedia. “Moltbook.”

Author

Sandy Carter

Sandy Carter is a globally recognized technology executive, author, and thought leader working at the intersection of artificial intelligence, blockchain, and digital identity. She has led and scaled multibillion-dollar business units at AWS and IBM and has held senior leadership roles across global technology organizations. Sandy is a Microsoft MSN AI Entrepreneur of the Year and a Fortuna AI and Innovation Leader of the Year. She serves as Chair of the Applied AI workgroup at The Digital Economist and is a Forbes contributor. Sandy is the author of the best selling book, AI First, Human Always, and the founder of Unstoppable Women of Web3 and AI, an initiative advancing women's leadership in emerging technologies worldwide.

Contributors

Arvinder (Singh) Kang

Arvinder Kang is a GovTech transformation leader with expertise at the intersection of artificial intelligence, public sector innovation, and ethical system design. He serves as Program Director at The Digital Economist and leads digital transformation initiatives at BCLC. Previously, as CTO of UrbanLogiq, Arvinder built and scaled data-driven platforms supporting policy development, infrastructure planning, and civic decision-making. He is a PhD candidate specializing in AI ethics and the evaluation of large language models in low-resource environments, with a focus on responsible deployment at scale. Arvinder regularly advises organizations and startups on AI governance, trust frameworks, and the practical implications of autonomous systems in regulated environments.

Priyanka Shrivastava, PhD

Priyanka Shrivastava, PhD, is Associate Dean and Professor at Hult International Business School, where she teaches and conducts research at the intersection of business strategy, innovation, and emerging technologies. A TEDx speaker, author, and business consultant, she works closely with startups, executives, and global organizations on leadership, transformation, and future-ready business models. Her research explores how technological change reshapes organizational decision-making, human capital development, and value creation in complex and uncertain environments. Priyanka is also an active startup mentor and thought leader, translating academic insight into practical frameworks for leaders navigating accelerated technological disruption.

Erika Twani

Erika Twani is an author, TEDx speaker, and nationally recognized leader in education, equity, and technology-driven transformation. She is the founder and CEO of the Learning One to One Foundation and has served as a Fortune 100 executive, board member, and senior executive fellow. Her work advances equitable access to education and prepares institutions for AI-enabled learning. She has advised leaders across government, education, and industry on scaling innovation responsibly while centering human outcomes. A sought-after keynote speaker, Erika brings a systems-level understanding of how technology, policy, and leadership intersect to shape long-term societal impact.

Nithin Singh Mohan

Nithin Singh Mohan is an AI and supercomputing leader at Hewlett Packard Enterprise, where he builds and scales advanced AI systems at supercomputing scale. He brings extensive experience spanning enterprise AI, high-performance computing, and startup innovation, including leadership roles within unicorn-stage companies. As a Senior Executive Fellow at The Digital Economist, Nithin contributes thought leadership on the future of AI infrastructure, agentic systems, and compute-driven innovation. His work bridges deep technical expertise with strategic insight, helping organizations translate large-scale AI capabilities into measurable business and economic impact.

About

The Digital Economist, headquartered in Washington, D.C. with offices at One World Trade Center in New York City, is the world's foremost think tank on innovation advancing a human-centered global economy through technology, policy, and systems change. We are an ecosystem of 40,000+ executives and senior leaders dedicated to creating the future we want to see—where digital technologies serve humanity and life.

We work closely with governments and multi-stakeholder organizations to change the game: how we create and measure value. With a clear focus on high-impact projects, we serve as partners of key global players in co-building the future through scientific research, strategic advisory, and venture build out.

We engage a global network to drive transformation across climate, finance, governance, and global development. Our practice areas include applied AI, sustainability, blockchain and digital assets, policy, governance, and healthcare. Publishing 75+ in-depth research papers annually, we operate at the intersection of emerging technologies, policy, and economic systems—supported by an up-and-coming venture studio focused on applying scientific research to today's most pressing socio-economic challenges.

CONTACT: INFO@THEDIGITALECONOMIST.COM