The Native Layer

Why AI agents interacting with software — and each other — is now a certainty: the evidence, the obstacles, and what to watch next.

Executive summary

This paper makes one claim and defends it: **software interaction is moving from human interfaces to machine-native interfaces — AI agents talking to services, data, and each other over purpose-built protocols — and this shift is now a certainty**, in the sense that matters for planning: the standards exist, the money rails exist, the operating systems are being rebuilt around it, the capability curves are compounding, and the world's most conservative buyers are already deploying at scale.

Five classes of proof support the claim: (1) fierce competitors standardized on shared agent protocols and surrendered them to neutral governance; (2) payment networks and the largest retailers shipped agent-native transaction rails; (3) the operating system and the cloud became agent hosts, and agent platforms now produce billion-dollar revenue lines; (4) agent capability doubles on a measurable curve while inference prices collapse; (5) the U.S. defense establishment — the hardest customer on earth — moved a million-plus users onto agent-building platforms.

Certainty of direction is not certainty of timeline, and it is not safety. The obstacles are real and mostly unsolved: prompt injection has no general fix, agent identity standards are still drafts, error rates compound across steps, most enterprise pilots still fail to show returns, and liability is being priced by a nervous insurance market. The paper closes with the signals worth watching next — the leading indicators that will tell you the future is arriving on, ahead of, or behind schedule.

1. The claim, precisely

For fifty years, software assumed a human at the point of interaction: screens, buttons, sessions, passwords. The machine-native model inverts that. An agent — a model with goals, tools, and memory — interacts with a service through an interface built for machines: a described tool it can call, a resource it can read, another agent it can message, a payment it can sign. The human moves up a level, from operator to delegator.

The claim is not that chat interfaces win, or that any particular vendor wins. It is that the *interface layer itself* is being rebuilt for machine consumers, the way it was once rebuilt for browsers and then for mobile. When interfaces change, everything downstream — security, identity, commerce, org charts — changes with them.

2. Proof class one: rivals standardized, then let go

The clearest signal that a protocol is infrastructure rather than product is when competitors adopt it and then give it away.

In November 2024 Anthropic released the Model Context Protocol (MCP), an open standard for connecting AI systems to tools and data. In March 2025, OpenAI — Anthropic's most direct rival — adopted it.[1] Through 2025, Google, Microsoft, Cursor, and VS Code followed. By December 2025, Anthropic donated MCP to the newly formed Agentic AI Foundation under the Linux Foundation, co-founded with Block and OpenAI, with AWS, Bloomberg, Cloudflare, Google, and Microsoft as platinum members — and more than 10,000 published MCP servers in the wild, spanning developer tools to Fortune 500 deployments.[2]

The same pattern repeated one layer up. Google's Agent2Agent protocol (A2A) — for agents discovering and messaging other agents — was donated to the Linux Foundation in June 2025 and within a year had over 150 supporting organizations and production deployments at Microsoft, AWS, Salesforce, SAP, and ServiceNow.[3]

This is the TCP/IP pattern, and it has one historical meaning: participants believe the layer is permanent. Companies do not hand their competitive moats to a foundation. They hand over the plumbing they expect to build on for decades.

3. Proof class two: money moves over it now

Commerce is the acid test of any interaction paradigm, because payments infrastructure is expensive, regulated, and only gets built for durable behavior.

In a nine-month window: OpenAI and Stripe launched agentic checkout inside ChatGPT (September 2025); Google and Shopify launched the Universal Commerce Protocol with Walmart, Target, Etsy, and Wayfair on board and endorsements from Amex, Best Buy, Home Depot, Mastercard, Stripe, and Visa (January 2026); Google's Agent Payments Protocol — cryptographically signed mandates proving what a user authorized an agent to buy — gathered 60+ partner organizations and moved to the FIDO Alliance for standardization (reported April 2026).[4][5] Visa and Mastercard each built agent-transaction rails of their own.[5]

Behavior followed the rails. Adobe's analytics show AI-referred traffic to U.S. retail sites up 393% year-over-year in Q1 2026 — and converting 42% *better* than ordinary traffic, a full reversal from a year earlier.[6] At the machine-to-machine frontier, Cloudflare and Coinbase's x402 protocol (HTTP-native micropayments that let agents pay for APIs and content per call) is reportedly serving on the order of a billion HTTP 402 responses daily across Cloudflare's network.[7]

Card networks, the biggest retailers, and both AI camps do not build settlement rails for a demo. They build them for a buyer that has already arrived.

4. Proof class three: the platforms rebuilt themselves around agents

The operating system is becoming an agent host. Windows 11 now ships native MCP support: an on-device registry of governed MCP servers, first-party servers for File Explorer and Settings, and sandboxed "agent workspaces" where agents operate under their own accounts.[8] Apple's beta releases show system-level MCP wiring through App Intents — unannounced, but visible in code.[9]

The cloud did the same. Amazon Bedrock AgentCore went GA in October 2025 and within six months spanned a dozen components — runtime, gateway, identity, memory, policy, evaluations, payments, registry — an operations stack, not a feature.[10] Google consolidated its enterprise agent offering into Gemini Enterprise; Microsoft shipped computer-using agents and an Agent Store inside Microsoft 365.[11]

And it produces revenue: Salesforce's Agentforce reached $1.2 billion in annualized revenue, growing 205% year-over-year, by May 2026 — the clearest evidence that enterprises pay real money for agents in production.[12] Gartner projects that by the end of 2026, up to 40% of enterprise applications will ship with task-specific agents embedded, up from under 5% in 2025.[13]

5. Proof class four: the curves all point the same direction

Three curves compound each other.

Capability. METR's measurements show the length of task agents can complete at 50% reliability doubling roughly every seven months since 2019 — and accelerating.[14] On OSWorld, a benchmark of real desktop computer work where humans score about 72%, agents went from roughly 7% at launch in 2024 to above the human baseline by mid-2026.[15] SWE-bench Verified, real software-engineering tasks, went from under 10% to over 70% in about a year.[15]

Cost. Inference prices are collapsing at rates between 9x and 900x per year depending on the task; the price of GPT-3.5-class capability fell 280-fold in

eighteen months.[16] Always-on background agents stop being a luxury when thinking becomes nearly free.

Efficiency. NVIDIA researchers argued in mid-2025 that small language models are the future of agentic AI: most agent subtasks do not need frontier models, small models are 10–30x cheaper to operate, and heterogeneous systems — small models routing to large ones only when needed — are the natural architecture.[17] Gartner expects task-specific small models to outnumber general-purpose LLM usage 3-to-1 by 2027.[18] This is the mainframe-to-PC inversion applied to intelligence: from renting access to owning the machine. For any organization whose data cannot leave its boundary, ownership is not a preference — it is the only compliant path.

6. Proof class five: the most conservative buyer already moved

Skeptics can dismiss startups and retail. It is harder to dismiss the Pentagon.

GenAI.mil launched in December 2025 and reached roughly 1.5 million defense users within six months; five of six military branches designated it their primary enterprise AI platform.[19] Its agent-builder tooling let DoD employees create on the order of 100,000 custom agents within weeks of release.[20] ChatGPT certified for controlled unclassified information at Impact Level 5 reaches more than 3 million personnel this month; Azure OpenAI is authorized through IL6.[21] The department's January 2026 strategy mandates deployment of new frontier models within 30 days of public release.[22]

When the institution whose failure modes are measured in lives — operating under ITAR, CUI, and classification law — deploys agents at population scale, the argument that agents are incompatible with serious environments is over. What remains is the engineering of boundaries.

7. The obstacles, honestly

Certainty of direction does not mean the road is clear. Six obstacles stand out, and none is fully solved.

Security: the unsolved core

Prompt injection — malicious instructions hidden in content an agent reads — has no general fix, only mitigations. 2025 produced a catalog of real incidents: a poisoned GitHub issue that caused agents to exfiltrate private repository data; an agent executing SQL embedded in a customer support ticket; a tenant-data leak that took a major SaaS vendor's MCP feature offline for two weeks; the first malicious MCP server found in the wild, quietly BCC'ing email to an attacker.[23] Microsoft's own documentation for Windows' agentic features warns that cross-prompt injection can lead to data exfiltration or malware installation — a vendor admitting the paradigm's defining attack class in its release notes.[24] Until progress here is structural, prudent enterprises will cap agent autonomy and keep humans on the approval path for consequential actions.

Identity and authorization: being built in real time

Agents act, so agents need identities, credentials, and least-privilege scopes — and the standards are still wet ink. Microsoft gave agents first-class directory identities (Entra Agent ID, 2025); Okta shipped per-user, per-action delegation tokens for agents (GA April 2026); the IETF and W3C both stood up agent-identity standardization efforts in spring 2026.[25] The gap between deployment and governance is measurable: one 2026 survey found 90% of executives confident in their AI visibility while half of knowledge workers use unapproved tools.[26]

Reliability: errors compound

A 99% success rate per step is a 60% failure rate across fifty steps. METR's data shows the task horizon at 80% reliability is four to five times shorter than at 50% — headline capability overstates dependable capability.[27] Worse, a systematic audit found many agent benchmarks have flaws that misestimate capability, in some cases by up to 100% in relative terms.[28] Enterprises should trust task-level evidence, not leaderboards.

Economics: the pilot graveyard

MIT's 2025 study found roughly 95% of enterprise GenAI pilots deliver no measurable P&L return; Gartner predicts over 40% of agentic AI projects will be canceled by the end of 2027 and coined "agent washing" for rebranded chatbots

and RPA.[29] Only 6% of companies say they fully trust agents with core processes.[30] The pattern in the failures is consistent: no governance, no data discipline, no ownership, no baseline against which to measure return.

Liability: priced, not resolved

An insurance market emerging around a risk is proof the risk is real. Lloyd's-backed underwriters began writing affirmative AI-liability policies in 2025; specialist policies now attach to agent audits at limits up to $50 million; meanwhile incumbent insurers are adding blanket AI exclusions at renewal, leaving agent errors uncovered by default.[31] Who answers for an agent's mistake — vendor, deployer, or model builder — remains largely untested law.

Fragmentation and regulatory drift

The tool layer consolidated on MCP, but agentic commerce has five-plus competing protocols, and identity standards are mid-draft — echoes of past standards wars that stalled adoption until convergence.[32] In Europe, the AI Act's high-risk obligations slipped roughly sixteen months (to late 2027 and 2028), easing near-term drag while extending uncertainty.[33] In regulated industries, data-boundary rules (CUI, ITAR, HIPAA) make architecture the gating decision: retrofitting compliance costs multiples of building it in.[34]

8. What it means for operators in regulated industries

The synthesis is not "wait." It is "build the boundary first."

- **Model-agnostic rails.** The customer side is codifying anti-lock-in; architectures should switch models on price, quality, and clearance without rework.
- **Owned intelligence where data cannot leave.** Small, private models inside the compliance boundary convert AI from a subscription expense into a balance-sheet asset — and are, for controlled data, the only fully compliant endpoint.
- **Read-only first, audit everything.** Agents that surface and recommend — with humans deciding — capture most of the value at a fraction of the risk, and generate the audit trail regulators and acquirers will ask for.
- **Govern identities now.** Every agent needs an owner, a scope, and a revocation path before it needs a use case.
- **Baseline before deployment.** The 95% pilot failure statistic is, in large part, a measurement failure. You cannot prove return against a baseline you never captured.

9. Watch for next

The leading indicators that will mark the next phase, roughly in expected order:

1. **Signed, attested agent registries.** Tool marketplaces adding cryptographic provenance — the "verified publisher" moment that made app stores safe for enterprises.
2. **OS agent workspaces on by default.** Windows agent sandboxes and Apple's App Intents work shipping enabled in enterprise fleets, not previews.
3. **Agent-to-agent transactions at volume.** Sustained growth in x402/AP2-style machine payments and the first large agent-pays-agent marketplaces.
4. **80%-reliability horizons crossing a workday.** When dependable (not median) autonomy covers eight hours, delegation economics flip for most back-office work.
5. **Commerce protocol convergence.** ACP, UCP, and AP2 interoperating or merging under FIDO/Linux Foundation governance — the Matter moment for agentic commerce.
6. **Agent governance as a budget line.** CISO organizations funding agent identity, policy, and observability as a category, with insurers pricing premiums to audit posture.
7. **A structural advance against prompt injection.** Architectural isolation of instructions from content — the single development that would most accelerate everything else.
8. **Regulated mission deployment.** Defense and healthcare agents crossing from administrative workflows into mission and clinical systems inside accredited boundaries.
9. **Small-model agents outnumbering frontier calls.** Telemetry confirming Gartner's 3-to-1 prediction — evidence the ownership model won the economics.
10. **The first major agent-caused loss and its settlement.** The case that establishes precedent for liability — painful, and clarifying, the way early cloud breaches were.

The direction is set. The pace will be decided by items one through seven. Watch those, not the demos.

Sources

1. TechCrunch, "OpenAI adopts rival Anthropic's standard," Mar 26, 2025 — techcrunch.com
2. Linux Foundation press release, Agentic AI Foundation formation (MCP donation; members; 10,000+ servers), Dec 9, 2025 — linuxfoundation.org
3. Linux Foundation, "A2A protocol surpasses 150 organizations," Apr 2026; Google Developers Blog, A2A donation, Jun 2025
4. Elogic, ChatGPT Instant Checkout / Agentic Commerce Protocol statistics, 2026; Google Developers Blog and Shopify Engineering, Universal Commerce Protocol, Jan 2026
5. Vellum, Google AP2 launch coverage, Sept 2025; PayRam protocol comparisons, 2026 (AP2-to-FIDO donation reported Apr 2026); Visa Intelligent Commerce Connect coverage, Apr 2026 (reported)
6. Adobe Analytics via Decrypt, AI-referred retail traffic +393% YoY, converting 42% better, Q1 2026
7. Cloudflare blog on x402; CoinDesk, May 5, 2026 (volume figures reported; single-source)
8. Microsoft Windows agentic developer documentation, 2025–2026 — developer.microsoft.com/windows/agentic
9. 9to5Mac, macOS 26.1 beta MCP/App Intents findings, Sep 22, 2025 (beta code; unannounced)
10. AWS What's New, Bedrock AgentCore GA, Oct 13, 2025, and component GAs through Mar 2026
11. The Register, Gemini Enterprise consolidation, Oct 9, 2025; Microsoft Copilot blog, computer-using agents GA, May 2026
12. Salesforce Q4 FY26 earnings release, Feb 25, 2026; CNBC, Q1 FY27 coverage, May 27, 2026
13. Gartner press release, Aug 26, 2025 (40% of enterprise apps with task-specific agents by end 2026)
14. METR, "Measuring AI ability to complete long tasks," Mar 2025, and time-horizons updates, 2026
15. OSWorld leaderboards (mid-2026, third-party; reported); vals.ai SWE-bench Verified tracker, 2026
16. Epoch AI, LLM inference price trends, 2025–2026; Stanford HAI AI Index 2025

17. NVIDIA Research, "Small Language Models are the Future of Agentic AI," arXiv 2506.02153, Jun 2025
18. Gartner small-model prediction (3x by 2027), 2026 coverage
19. War.gov GenAI.mil launch release, Dec 2025; DefenseScoop, usage and branch adoption, Feb–Jun 2026
20. DefenseScoop, "Pentagon uses GenAI.mil to create agents" (~100,000 agents), Apr 23, 2026
21. Nextgov/FCW, ChatGPT at IL5 on GenAI.mil, Jun 2026; Microsoft Azure Government blog, Azure OpenAI IL6 authorization
22. Department AI strategy, Jan 2026 — media.defense.gov; Holland & Knight analysis, Feb 2026
23. Docker "MCP Horror Stories" series, 2025; Checkmarx, MCP security incidents and controls, 2025–2026
24. Microsoft Support, "Experimental agentic features" (XPIA warning), Nov 2025
25. Microsoft Entra Agent ID announcement, May 2025; Okta for AI Agents GA, Apr 30, 2026; IETF draft-klrc-aiagent-auth, Mar 2026; W3C Agent Identity Registry CG proposal, Apr 2026
26. Okta, "AI agents at work" 2026 survey
27. METR time-horizon data; "Is there a half-life for the success rates of AI agents?" arXiv 2505.05115, May 2025
28. Zhu et al., "Establishing Best Practices for Building Rigorous Agentic Benchmarks," 2025 — openreview.net
29. Fortune on MIT "GenAI Divide" report, Aug 18, 2025; Gartner press release on agentic project cancellations and "agent washing," Jun 25, 2025
30. Fortune / Harvard Business Review survey coverage, Dec 9, 2025 (6% full trust)
31. ACM News, "AI liability insurance arrives," 2025; AIUC audit-linked policies, 2026; ABA, AI insurance policy-gap analysis, Fall 2025
32. PayRam and Orium, agentic payments protocol landscape, 2026
33. Gibson Dunn, EU AI Act Digital Omnibus (high-risk deadlines postponed), May–Jun 2026
34. Kiteworks, AI compliance in healthcare (retrofit cost multiples), 2026

Figures marked "reported" derive from single or secondary sources and are presented as reported rather than independently verified. All other claims trace to the primary sources listed above. Compiled July 2026.