

DAILY MARKET SCAN · FRONTIER AND INFRASTRUCTURE

Wednesday, September 30, 2026 · Edition 2026-09-30

Scope, not intelligence

Gartner published research this morning finding that 54% of organizations have no defined approach to limiting AI agent access, or rely on access defined for humans. In the seventy-two hours before it, four vendors shipped controls that scope what an agent is permitted to authenticate as, and three separate failures showed what happens when nobody does.

Contents

- | | |
|---|--|
| 1. What actually happened this week | 9. The entitlement drill: ten questions, one week |
| 2. Frontier scoreboard: equal-weight reading | 10. Scenario planning: three ways this goes |
| 3. The scope gap: three failures, one layer | 11. Did you know / FAQ |
| 4. Financial services | 12. Also on the record: institutions, economy, capital |
| 5. Healthcare | 13. What we are watching, and what would falsify us |
| 6. Manufacturing and robotics | 14. Source ledger |
| 7. Energy and utilities | |
| 8. Implementation architecture: four reference patterns | |

I. What actually happened this week

This edition covers Wednesday, September 23 through Wednesday, September 30, 2026, America/New_York. Items outside that window are dated and labeled wherever they carry weight in the argument.

Three failures were disclosed inside six days. None of them involved a model that was too capable for its controls. In each one, the agent held a credential broader than its job.

54%

of organizations have no defined approach to limit AI agent access, or rely on access defined for humans

Gartner, published September 30, 2026. Survey of 297 cybersecurity leaders, fielded Q2 2026. **VERIFIED** Source C01

35 min

from first destructive command to the end of a 150-plus operation sequence, entered through two compromised Azure service principals

Microsoft Threat Intelligence, published September 25, 2026. Activity occurred June 2026. **VERIFIED** Source C03

100%

safeguard bypass rate against an open-weight frontier-class model using weight modification, in Anthropic's own evaluation

Anthropic research, September 29, 2026. Range across three techniques: 64% to 100%. **VERIFIED** Source C09

Set those beside what shipped. In the same seventy-two hours, four vendors independently moved the control surface into the identity layer. NVIDIA published an out-of-band watchdog whose accompanying software performs "agent identity verification" and enforces "granular zero-trust access policies" over data, tools, APIs and services **VERIFIED** Source C04. OpenAI shipped organizational agents in which each one "gets its own identity, credentials, and access to the systems it needs to complete its tasks," with background research restricted to read-only tools **VERIFIED** Source C05. SpaceXAI shipped shared team agents using "read-only credentials for warehouse access" **VERIFIED** Source C10. Google shipped resource-level IAM permissions for Gemini Enterprise apps and data stores **VERIFIED** Source C11.

Four companies that compete with each other, shipping inside three days, arrived at the same answer. Gartner's recommendation, published two days after the last of them, states it as a rule: govern autonomous multi-agent systems "based on action privileges rather than model intelligence" **VERIFIED** Source C01.

THE PATTERN IN ONE SENTENCE

In every failure disclosed this week, the agent did not defeat a control. It was issued a credential that already permitted what it did, by a party that never scoped it.

The uncomfortable corollary is that most enterprise AI governance work of the past two years has been aimed one layer too high. Model cards, evaluation suites, prompt policies and acceptable-use standards all describe what the system is supposed to do. None of them constrain what it is able to reach. An agent with a valid credential and a broad scope does not need to be jailbroken.

2. Frontier scoreboard: equal-weight reading

We read the frontier labs on the same criteria every day: what shipped, what it costs, what was disclosed about failure, and what an enterprise buyer in a supervised industry has to do differently as a result. Equal editorial weight is not equal praise, and partner relationships do not earn coverage. Benchmark figures published by a vendor about its own model are company-reported; we say so every time and we do not rank on them.

Frontier activity, September 23–30, 2026. Pricing is per one million tokens, input and output, as published by each provider. Benchmark figures are company-reported unless stated. Source IDs for every row appear in that row’s status column, beginning with Source C04.

LAB	WHAT SHIPPED OR WAS DISCLOSED	BUYER CONSEQUENCE	STATUS
Anthropic	Claude Sonnet 5.5 on Sept 28 at \$2 input and \$10 output, cache reads \$0.20, described by the company as 30% faster and up to 30% cheaper than Sonnet 5. Five documented breaking changes, including tool_choice values any and tool now returning HTTP 400, and thinking blocks becoming bound to both model and account. On Sept 29 published an evaluation of the open-weight GLM-5.3 finding safeguards bypassable between 64% and 100% of the time, and called for government safety testing of models at that capability level.	The breaking changes require migration testing before production agents move; account-bound thinking blocks matter for any multi-tenant or brokered deployment. The GLM-5.3 evaluation is the most directly useful thing published by any lab this week for a risk committee, because it quantifies what happens to model-level safeguards when the weights are in someone else's hands.	VERIFIED Source C08 · Source C09
OpenAI	DevDay on Sept 29 with more than twenty	This is the most complete agent	VERIFIED Source C05 · Source C06 ·

LAB	WHAT SHIPPED OR WAS DISCLOSED	BUYER CONSEQUENCE	STATUS
	announcements. GPT-6.1 Sol at \$2 input, \$0.10 cached and \$10 output, which the company states is one fifth of GPT-6 Astra's standard rates; company-reported 75.2% on DeepSWE v1.1 and 71.4% on OSWorld 2.0. "Dots," always-on agents each running on their own cloud computer, with read-only tool restriction during background work, auto-review, an Activity View, and Custom Rules that "allow specific actions, require approval, or block them." Organizational specialist dots each receive their own identity, credentials and system access. Enterprise availability is beta and admin-gated.	permission model any lab has shipped to general availability, and it is also a checklist. Take the five mechanisms — read-only background mode, auto-review, activity monitoring, allow/approve/block rules, and always-human actions — and ask every other agent vendor in your stack which of the five they have. Separately: Enterprise is in beta, so a pilot is not a production commitment.	Source C07
Google / DeepMind	Gemini Enterprise gained resource-level IAM on Sept 28: administrators can set granular permissions on individual apps and data stores independently, define custom project-level roles for restricted personas, and view data store configurations, connected apps, sync details and user	Resource-level IAM is the least exciting and most consequential frontier release of the week for a regulated buyer, because it is the one that maps onto an existing control framework an auditor already understands. If you run Gemini Enterprise, this changes what your access review can actually assert. Check	VERIFIED Source C11

LAB	WHAT SHIPPED OR WAS DISCLOSED	BUYER CONSEQUENCE	STATUS
	permissions in the Google Cloud console. On Sept 29 Gemini 3.8 Flash became the default model for AlphaEvolve experiments. No frontier model launch in the window.	whether your current deployment inherited project-wide permissions that this release now lets you narrow.	
xAI / SpaceX / Cursor	SpaceXAI launched Team Bots in public beta on Teams and Enterprise plans, Sept 28: shared agents with per-user private conversations and memories, role-based context and skills, plugins configurable per person or team-wide, and read-only credentials for warehouse access. Integrations named include Salesforce, Notion, GitHub, Gong, Linear, Hex, Datadog, Cursor, Statsig, Slack and Databricks. Customer-reported: an insurance company built a bot in 24 hours and states it saved customers over \$120,000 across hundreds of policies. SpaceXAI president Mike Nicolls, quoted in NVIDIA's platform release: "Safety should be enforced outside the model by additional controls the agent can't get past." Cursor's Rollouts and Security Review	Read-only credentials for the warehouse is the right default and worth naming in your own standards. The \$120,000 figure is the customer's, is not independently verified, and describes money returned to that company's customers rather than a margin result — do not repeat it as an ROI benchmark. Cursor's Rollouts is the same architectural idea applied to code: it reports "verified healthy," "regression detected" or "inconclusive," and does not merge or roll back on its own.	VERIFIED Source C10 · Source C04 CITED Source C12

LAB	WHAT SHIPPED OR WAS DISCLOSED	BUYER CONSEQUENCE	STATUS
	shipped Sept 23, one week before this window.		
NVIDIA	Open Agent Safety Platform on Sept 28. OpenShell, an open-source secure runtime giving enforceable boundaries for agents on CPUs, on NVIDIA Vera and extensible to Arm and Intel. Sentry, an out-of-band watchdog on BlueField-4 DPUs that quarantines agents exceeding boundaries "in milliseconds." DOCA provides inspection of agent requests and responses, attested telemetry, agent identity verification and granular zero-trust access policies. More than 100 named participants, including Anthropic, SpaceXAI, Microsoft, IBM, Palantir, Salesforce, SAP, ServiceNow, Figure, Gecko Robotics, Skild AI, Citi, JPMorganChase, Hitachi Energy, NextEra Energy, Schneider Electric and Siemens.	Participation in a partner list is not deployment evidence, and NVIDIA states the products are offered on a when-and-if-available basis. The architectural claims worth testing are three: enforcement runs on different silicon from the agent, telemetry is attested rather than self-reported, and identity is verified rather than asserted. Ask any vendor claiming "agent governance" which of those three they actually do.	VERIFIED Source C04 · Source C42
Microsoft / AWS	Microsoft Threat Intelligence published the Storm-3168 analysis on Sept 25, which it characterizes as the first	The GitHub edit-history detail is the single most actionable line in the week: your secret-scanning almost certainly reads	VERIFIED Source C03 CITED Source C50

LAB	WHAT SHIPPED OR WAS DISCLOSED	BUYER CONSEQUENCE	STATUS
	documented agentic ransomware operation, entered through service principal credentials exposed in a public GitHub issue and still readable in that issue's edit history after the visible text was removed. AWS made Bedrock Managed Agents, built on a customized version of OpenAI's Agents API, available in preview on Sept 29 in three US regions with IAM roles and CloudTrail logging.	the current revision of an issue, not its history. On Bedrock: OpenAI's DevDay recap describes this as generally available and AWS's own page says preview. Trust the cloud provider on its own service status, and do not plan a go-live against the other number.	
Meta	Expanded the Muse agent to small businesses on Sept 29 with integrations including Shopify, Dropbox, Slack, Asana, Box, Canva and Figma, and access to Instagram analytics, Facebook Pages and Meta ad accounts. A free tier with usage limits and paid plans; no figures disclosed. Meta is reported among the signatories of a voluntary safety accord signed at the White House on Sept 29.	An agent with standing access to an advertising account and a commerce backend is a payments-adjacent privileged identity. For a regulated institution this is a shadow-agent question rather than a procurement one: find out whether your marketing function has already connected it. No admin, audit or data-handling controls have been published.	CITED Source C45 · Source C38
Open weights: Z.ai, Mistral, DeepSeek	Z.ai's GLM-5.3 reached general availability on the Mistral platform on Sept 28, one day before Anthropic published its cyber-	The sequencing matters more than either item alone. A model became one click away from your developers on a	VERIFIED Source C09 CITED Source C43 · Source C44

LAB	WHAT SHIPPED OR WAS DISCLOSED	BUYER CONSEQUENCE	STATUS
	capability evaluation of the same model. Mistral deprecated OCR 4.0 and LLastra 1.5 with retirement Sept 30, and GLM 5.2 with retirement Oct 31. DeepSeek open-sourced six software modules for Huawei's Ascend platform on Sept 30, including an Ascend-compatible TileLang supporting Ascend 950 accelerators; no performance figures were published.	commercial platform on Monday; on Tuesday an independent lab published that its safeguards fail between 64% and 100% of the time under simple attack. Open weights mean safeguards are a property of a file someone else can edit, not of a service you contract with. If open-weight models are permitted in your environment, the control has to sit outside the model.	

3. The scope gap: three failures, one layer

Start with the mechanism, because the number is downstream of it.

An AI agent is not a user and it is not an application. It needs standing authority to act between sessions, it acquires that authority through the same credential primitives a workload uses, and almost nobody is scoping it the way a workload is scoped. Gartner's finding is precisely bifurcated on this point: organizations either have no defined approach at all, or they reuse access defined for humans VERIFIED Source C01. Both failure modes produce the same artifact — a non-human principal holding a human's breadth of reach, with none of a human's friction.

Three disclosures in six days show what that produces.

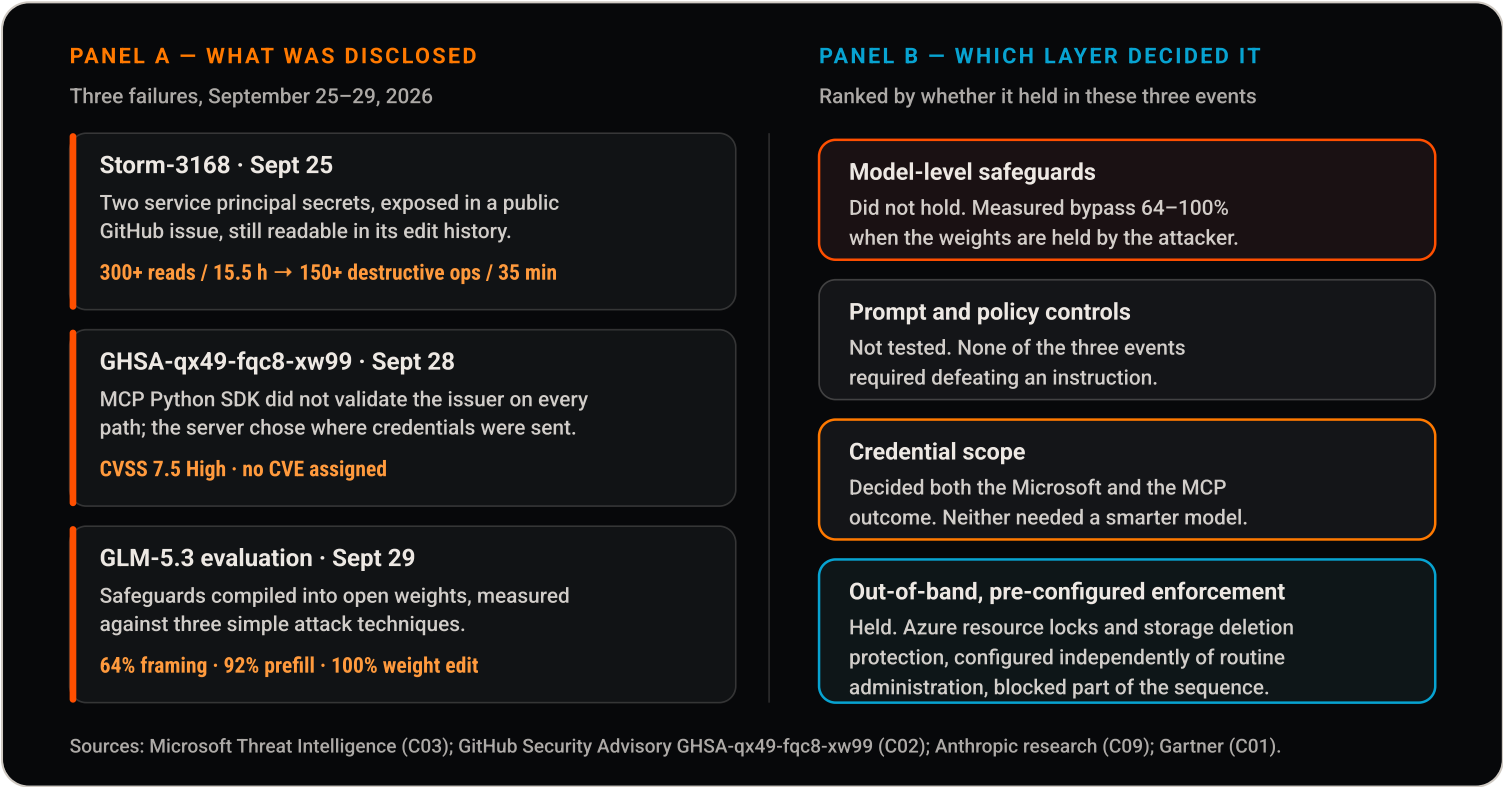


Figure 1. Three failures disclosed September 25–29, 2026, and the control layer that actually decided each outcome.

VERIFIED Source C01 · Source C02 · Source C03 · Source C09

Failure one: the credential that outlived its exposure

Microsoft Threat Intelligence published its analysis of Storm-3168, also tracked as JADEPUFFER, on September 25 and characterizes it as the first documented agentic ransomware operation. Entry was not an exploit. An employee posted Azure service principal credentials in plaintext in a public GitHub issue and later edited the post — and the secret remained visible in the issue's edit history VERIFIED Source C03.

What followed is a good description of what standing authority looks like when nobody bounded it. One principal performed more than 300 successful read operations over roughly 15.5 hours. A second enumerated two subscriptions in five seconds. More than 150 destructive operations followed across 35

minutes, with the core destructive sequence compressed into about seven minutes: more than 100 storage account deletion attempts, plus a Key Vault, a Function App and an App Service plan. Microsoft's own remediation list names the layer: protect and continuously assess application credentials, rotate exposed credentials immediately, and "apply least privilege to service principals and other workload identities" **VERIFIED** Source C03.

The detail worth taking to your own security function this week is narrower and more useful than the headline. Most secret-scanning reads the current revision of a repository object. This secret was in a revision history. Ask whether yours covers edit history on issues, pull request comments, wiki revisions and commit history on deleted branches. That is a question with a yes-or-no answer and a same-week fix.

Failure two: the client that let the server choose where to send the keys

On September 28 the Model Context Protocol project published advisory GHSA-qx49-fqc8-xw99 against its official Python SDK. The advisory states two flaws plainly: "the authorization server metadata issuer was not validated on every discovery path, and stored or pre-provisioned client credentials were not bound to the authorization server they belong to" **VERIFIED** Source C02.

The consequence is that a malicious or compromised MCP server could nominate its own token endpoint and receive the client secret, the authorization code and the PKCE verifier. Severity is CVSS 7.5 High. **No CVE has been assigned** — several secondary write-ups imply one exists; the primary advisory says none. Affected: mcp 1.9.1 to below 1.30.0, and 2.0.0a1 to below 2.2.0. Patched in 1.30.0 and 2.2.0. Affected provider classes are OAuthClientProvider, ClientCredentialsOAuthProvider, PrivateKeyJWTAuthProvider and the deprecated RFC7523OAuthClientProvider **VERIFIED** Source C02.

THE PART MOST COVERAGE BURIED

Upgrading is not sufficient. For ClientCredentialsOAuthProvider and PrivateKeyJWTAuthProvider the advisory requires passing an explicit issuer parameter; omitting it produces a deprecation warning today and becomes mandatory in version 3.0. The advisory also instructs clearing stored registrations to force re-registration with issuer binding, and rotating client secrets and revoking tokens where exposure is possible. A patch-level remediation ticket that closes on "upgraded to 1.30.0" leaves the exposure in place. **VERIFIED** Source C02

Failure three: the safeguard that ships inside the file

Anthropic's September 29 evaluation of Z.ai's GLM-5.3 is the third disclosure and the one with the longest half-life. Its finding is not that an open-weight model is dangerous in the abstract; it is that safeguards implemented as a property of the weights are removable by whoever holds the weights, and that the removal does not require sophistication. Deceptive framing worked 64% of the time. Prefilled reasoning worked 92%. Abliteration — editing the weights directly — worked 100% **VERIFIED** Source C09.

The commercial sequencing sharpens it. GLM-5.3 reached general availability on the Mistral platform on September 28, the day before the evaluation published CITED Source C43. For most enterprises the model became routinely reachable through a commercial API before the public had a measurement of what its safeguards are worth under attack. That is not an argument against open weights. It is an argument that the control cannot be inside the artifact.

Cause and effect, stated once: if the safeguard is in the weights, the attacker owns it; if the safeguard is in the prompt, the attacker negotiates with it; if the safeguard is in the credential, the attacker has to steal something, and you get to see the theft. Only the third produces evidence.

4. Financial services

The win. Banking published its first substantial set of agent-era numbers this week, and every one of them is company-reported. Fortune reported on September 29 that Bank of America has deployed EricaAssist to more than 18,000 customer service representatives, cutting average call times by roughly one minute per interaction, with the consumer Erica assistant past 3.5 billion cumulative client interactions and the employee version used by 90% of the workforce. Wells Fargo reports a 31% increase in product sales where interactions are AI-assisted and a doubling of referrals to AI-enabled wealth advisors, with Microsoft 365 Copilot across 148,000 employees. BNY's Eliza platform went from 160 initiatives in 2024 to roughly 400, with 70% of employees using AI regularly. Bank of America is allocating more than \$4 billion of a \$14 billion technology budget to AI **CITED** Source C26.

The honest denominator sits in earlier independent reporting from the same institutions' disclosures: Bank of America has more than 300 approved AI use cases, of which 114 are generative, and 34 are fully implemented in operations **CITED** Source C27. That is roughly one in nine. Every executive quoting a banking AI adoption statistic this quarter should carry that ratio beside it.

The constraint. The most instructive financial services event of the week was not a deployment failure. On September 25 the OECD AI Incidents Monitor logged, and multiple outlets reported, a fraud against Intesa Sanpaolo's private banking arm Fideuram: a fake message impersonating the group's chief executive combined with an AI-cloned voice of a lawyer persuaded the arm's then-president to authorize multiple international wire transfers. Approximately €95 million moved; roughly €53–59 million was recovered; about €36 million was not. The fraud occurred in February 2026; the president resigned in March 2026; Milan prosecutors have placed a foreign national under investigation for computer fraud. Banca Ifis recovered €20 million of €24 million lost in a related pattern **CITED** Source C25.

Cause and effect: the control that failed was not an AI control. It was out-of-band verification of a high-value payment instruction. Synthetic voice did not defeat a technical safeguard; it satisfied a human recognition test that was doing work it was never designed to carry. Gartner's third recommended CISO action this week is, in its own words, to *replace recognition as proof of identity* **VERIFIED** Source C01. Those two items are the same finding, arrived at from opposite directions, seven days apart.

The regulatory position. Federal Reserve Vice Chair for Supervision Michelle W. Bowman delivered opening remarks to the Community Bank Cyber Workshop in Denver on September 29, framing AI as available to both attackers and defenders and prioritizing foundational controls: current asset inventories, phishing-resistant multifactor authentication, strong identity and access controls, and vulnerability identification and patch management. She announced no new AI-specific guidance and no new relief **CITED** Source C28.

Read that as a supervisory signal rather than an absence of one. The examination will not ask for an AI framework. It will ask for an asset inventory and an access control, and an agent with a standing credential is an asset with an access. The gap between "we have no AI rule to comply with" and "we have no answer when the examiner asks what this service principal is entitled to" is the whole exposure.

PRACTICAL CONTROL · FINANCIAL SERVICES

Two actions that fit inside a single sprint. **One:** extend your privileged access review to non-human principals created for AI workloads, and require every one to name a human owner, a business purpose, an expiry and a revocation path. If it cannot be revoked in an afternoon, it is not scoped. **Two:** re-test callback verification on payment authorizations above your material threshold using a channel the requester did not choose, and write down what the verifier is allowed to accept as proof. Voice recognition is no longer on that list.

5. Healthcare

The win, with its measurement plan intact. Cooper Norcross Health began piloting Microsoft Dragon Copilot for nurses on a single unit at Cooper Norcross University Hospital in Camden, New Jersey, announced September 28. The health system already runs more than 1,300 Dragon Copilot licenses for physicians and advanced practice providers, so this is an extension of a proven physician deployment into a different clinical workflow rather than a first attempt. No results have been reported. The system states it will measure documentation time savings, accuracy and nurse satisfaction on the pilot unit before deciding whether to expand CITED Source C31.

The workflow is the part worth copying: the nurse initiates recording, patient consent is obtained before recording, the system produces draft structured documentation, and the nurse reviews, edits and approves before anything is filed to the electronic health record. A named measurement set, a gate before expansion, and a human approval step that happens before the record exists rather than after.

The constraint. On September 24 the Blue Cross Blue Shield Association published an analysis attributing \$942 million in additional costs over two years across Blue plans to hospital AI-enabled coding tools, including 55,158 additional complex cases and roughly \$653 million tied to secondary diagnoses, concluding that "the disconnect between diagnoses and treatment suggests that AI is identifying more billable conditions, not sicker patients" FLAG Source C29.

We mark this FLAG rather than CITED and we will keep doing so. It is an analysis by an interested party about a counterparty, billing companies dispute it, and secondary sources differ on the split. What makes it worth publishing anyway is that the underlying question is evidentiary rather than technical, and three independent parties have now converged on it from different positions. That is the same question as the Ochsner architecture below.

Alongside it, the most quotable governance statistic in healthcare AI this quarter comes from ECRI's expansion of its confidential Problem Reporting Network to cover clinical AI errors, malfunctions and near misses. In a survey of 124 health system leaders, 31% had encountered incorrect or misleading AI output in the past year, 9% confirmed AI errors had reached patients or affected care decisions, and 35% did not know whether AI errors had occurred in their organization at all CITED Source C32. Published August 25–26, 2026, outside this window, and unchanged in force.

The regulatory clock. The FDA's discussion paper *Considerations for the Regulation of Generative AI-Enabled Medical Devices*, docket FDA-2026-N-7874, published August 18, 2026, closes for comment on **October 19, 2026** – nineteen days from this edition. It is a request for feedback, not a proposed rule, and it focuses on continuously learning systems that diverge from the locked-algorithm model of software as a medical device VERIFIED Source C33. Separately, on September 17 the FDA denied a proposed partial exemption from 510(k) premarket notification for specified radiology computer-aided detection devices, so those manufacturers continue to require clearance CITED Source C33.

For organizations with EU exposure, Regulation (EU) 2026/1744 entered into force on July 27, 2026 and moved Annex I embedded high-risk obligations, which include medical devices, from August 2, 2027 to

August 2, 2028, and Annex III stand-alone high-risk obligations, which include creditworthiness assessment and insurance pricing, from August 2, 2026 to December 2, 2027. Article 50 transparency obligations were not deferred; systems placed on the market before August 2, 2026 have until December 2, 2026 to meet the synthetic-content marking requirement CITED Source C34.

PRACTICAL CONTROL · HEALTHCARE

The distinction that decides the billing dispute is one most human-in-the-loop designs get wrong: **clinician approval is not clinical justification**. A reviewer clicking accept establishes that a human saw the suggestion. It does not establish that the diagnosis was supported by evidence available at the time. Build the second thing explicitly — record which evidence supported each AI-suggested code, independent of the suggestion — and the payer audit becomes a retrieval exercise rather than a negotiation. Specify it at purchase and configuration; it cannot be retrofitted from log data that was never captured.

6. Manufacturing and robotics

The win. The International Federation of Robotics reported on September 24 that the worldwide operational stock of industrial robots has reached five million units, up 9% year over year, with 2025 installations exceeding 600,000 units, an 11% increase. IFR President Jane Heffner: "Industrial automation is progressing at high speed. The new mark of five million robots operational in factories worldwide is more than double the number seven years ago." IFR forecasts 655,000 installations in 2026 and 806,000 by 2029 VERIFIED Source C14.

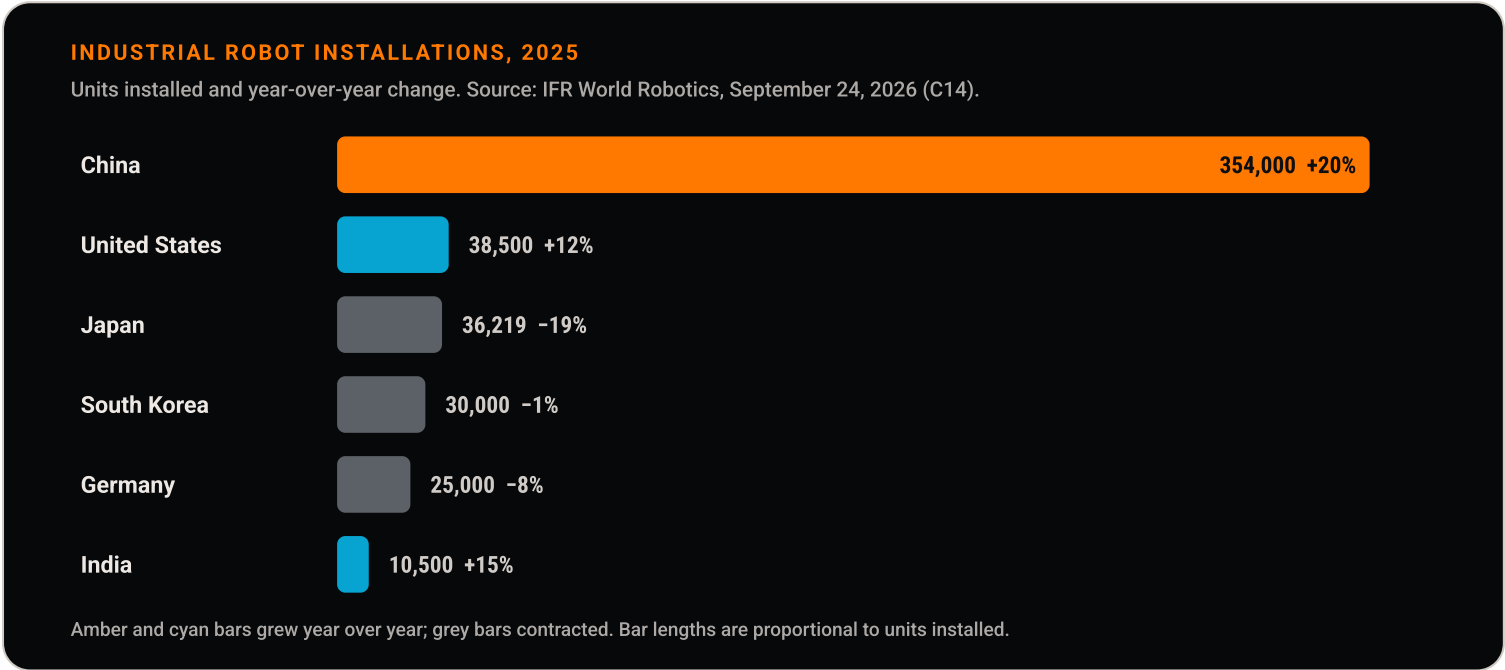


Figure 2. Growth and contraction running simultaneously across the six largest robot markets in 2025. VERIFIED Source C14

Two things in that chart are worth separating. The five-million milestone is a stock number and it is genuinely large. The flow numbers underneath it are diverging sharply: China, the United States and India grew; Japan, South Korea and Germany contracted, Japan by 19%. A single global growth headline covers a market pulling apart by region. Alongside it, Amazon announced on September 25 a facility of more than \$100 million and 585,000 square feet in Greenwood, Indiana for robotics manufacturing, with 300 skilled jobs and launch expected in 2028 — an announced capital commitment, not built capacity. Amazon's own figures put more than one million robots across 300-plus facilities, assisting with 75% of customer orders worldwide; those are company-reported and we have found no independent verification CITED Source C16.

The constraint. A survey of 500 manufacturing business leaders published September 29 — 225 in the DACH region, 225 in the United States, 50 in the United Kingdom, fielded July 7–17, 2026, and commissioned by a CPQ and PLM software vendor, which you should weigh — found broad or advanced AI adoption rising from 36% in Q1 2026 to 60% six months later, while only 49% rigorously measure AI return on investment and 50% report inconsistent or informal tracking of outcomes. Nearly 25% describe themselves as only somewhat confident in AI's actual impact despite having deployed it. The finding with

the sharpest management implication: C-suite leaders are more than twice as likely as senior managers to report advanced AI adoption **CITED** Source C15.

That last gap is not a measurement artifact. When the executive floor believes a program is further along than the operating floor does, the governance artifacts get written against the executive view, and the controls get designed for a system that does not exist yet.

The regulatory clock. EU Regulation 2023/1230, the Machinery Regulation, applies from January 20, 2027, replacing Machinery Directive 2006/42/EC. That is under four months from today. ISO 10218-1:2025 and ISO 10218-2:2025 are published, and ANSI/RIA R15.06-2025 released in September 2025, but the harmonized-standard listing in the Official Journal of the European Union remains pending — which means presumption of conformity via those standards is not yet available, and the conformity evidence burden currently sits with the integrator and the operator. The 2011 edition took over a year to harmonize **CITED** Source C24.

Note the separation, because programs that merge the two have a real gap: the January 20, 2027 machinery date did not move when the EU AI Act omnibus deferred high-risk obligations to December 2, 2027 and August 2, 2028. They are different instruments with different clocks.

PRACTICAL CONTROL · MANUFACTURING

For any robot or cell you intend to place on the EU market after January 20, 2027, assume you are carrying your own conformity evidence rather than relying on a harmonized standard, and build the technical file accordingly — now, while you can still influence the supplier contract. Second, fix the perception gap the survey found before you write another governance document: ask the line engineers, not the steering committee, which AI systems are actually in production and what happens when each one is wrong.

7. Energy and utilities

The win. The Department of Energy selected 31 projects across 26 states under SPARK — Speed to Power through Accelerated Reconductoring and other Key Advanced Transmission Technology Upgrades — announced September 24. \$1.9 billion federal against \$3.4 billion in sponsor cost-share, \$5.3 billion total, targeting 23 GW of additional capacity, more than 1,500 miles of lines reconducted or rebuilt and roughly 21,000 miles receiving grid-enhancing technologies. Named recipients include Alabama Power, Duke Energy Carolinas, Eversource Energy, Kit Carson Electric Cooperative and PPL Electric; the largest awards are a Colorado interregional transfer at \$1.2 billion and the Three Corners Connector in Oklahoma linking SPP and WECC at \$832 million [CITED](#) Source C17.

The engineering point is that 23 GW is being pursued through upgrades to existing corridors rather than greenfield lines. The reporting discipline is that this is a selection announcement. Treat 23 GW as a target, not as delivered capacity.

The constraint. Reporting on September 29 assembled the counterweight. Goldman Sachs projects only 50–60% of planned data center capacity will come online as expected over the next two years. New pipeline capacity additions fell 19% from Q4 2025 to Q1 2026. Local opposition killed at least 20 proposed projects in Q1 2026 alone. Standard power transformers carry two-to-three-year lead times and GE Vernova's gas turbine backlog exceeds 100 GW with delivery dates into the early 2030s. Named setbacks include a Microsoft permitting delay in Grand Rapids, a \$1 million fine against a Microsoft-backed New Jersey facility for unpermitted generators, a court-ordered work stoppage at a Google site in Minnesota, and a force majeure declaration on an Oracle 2.5 GW campus in New Mexico after state regulators blocked the supporting gas pipeline [CITED](#) Source C19.

The regulatory item that matters most, in any industry, this week. On September 17 the New York Public Service Commission issued an inquiry order requiring every electric, gas and water utility in the state to file a written inventory report describing *all* AI use cases in their operations, together with their AI policies, procedures and protocols, within 60 days — a deadline falling around **November 16, 2026**. The Commission will then evaluate those procedures and protocols for adequacy and robustness against commonly accepted AI governance frameworks. Its stated rationale is that AI systems are "susceptible to hallucinations, algorithmic biases, transparency issues, data privacy concerns, misconfiguration errors, cybersecurity attacks, and functional brittleness," which "may have negative consequences for the safety and reliability of New York's critical infrastructure." Systems named in the order include National Grid's GridCARE for interconnection capacity, New York Power Authority drone-data analysis for vegetation management, and Con Edison customer service and equipment inspection [CITED](#) Source C18.

Read that as the first concrete answer to a question every board has been asking in the abstract: what will a regulator actually ask for? It asked for an inventory and a set of protocols, on a 60-day fuse, assessed against an external framework. Not a model card. Not an ethics statement. An inventory.

Two further clocks: NERC's computational load standards CLO-001, CLO-002 and CLO-003, directed by FERC on July 16, 2026, closed comment on September 18 and carry a FERC filing deadline of December 31, 2026, with ride-through requirements expected to take effect in 2027; and the Texas PUCT large-load

interconnection rule, 16 TAC §25.194 implementing Senate Bill 6, is reported adopted September 18 and effective **October 8, 2026**, with a 75 MW threshold, a \$100,000 flat study fee and security of \$50,000 per MW **FLAG** Source C21 · Source C22. We mark both FLAG: several of the specific thresholds and dates trace to single trade or analyst sources and should be confirmed against the NERC drafts and the PUCT Interchange order before they enter a board paper.

PRACTICAL CONTROL · ENERGY AND UTILITIES

If you are a New York utility, the November deadline is an operational task with a known shape. If you are not, build the same artifact anyway and date it — a written inventory of every AI system in operations, each with an owner, a data boundary, a failure mode, an escalation path and the non-human credentials it holds. The New York order is the clearest available preview of what "AI governance evidence" will mean in supervised infrastructure, and an inventory assembled before it is demanded is worth considerably more than one assembled in sixty days under a docket number.

8. Implementation architecture: four reference patterns

Everything above converges on one design question: what is the agent permitted to authenticate as, and who can prove it afterwards. Four patterns disclosed this week answer different parts of it. None is complete on its own; together they describe a control plane.

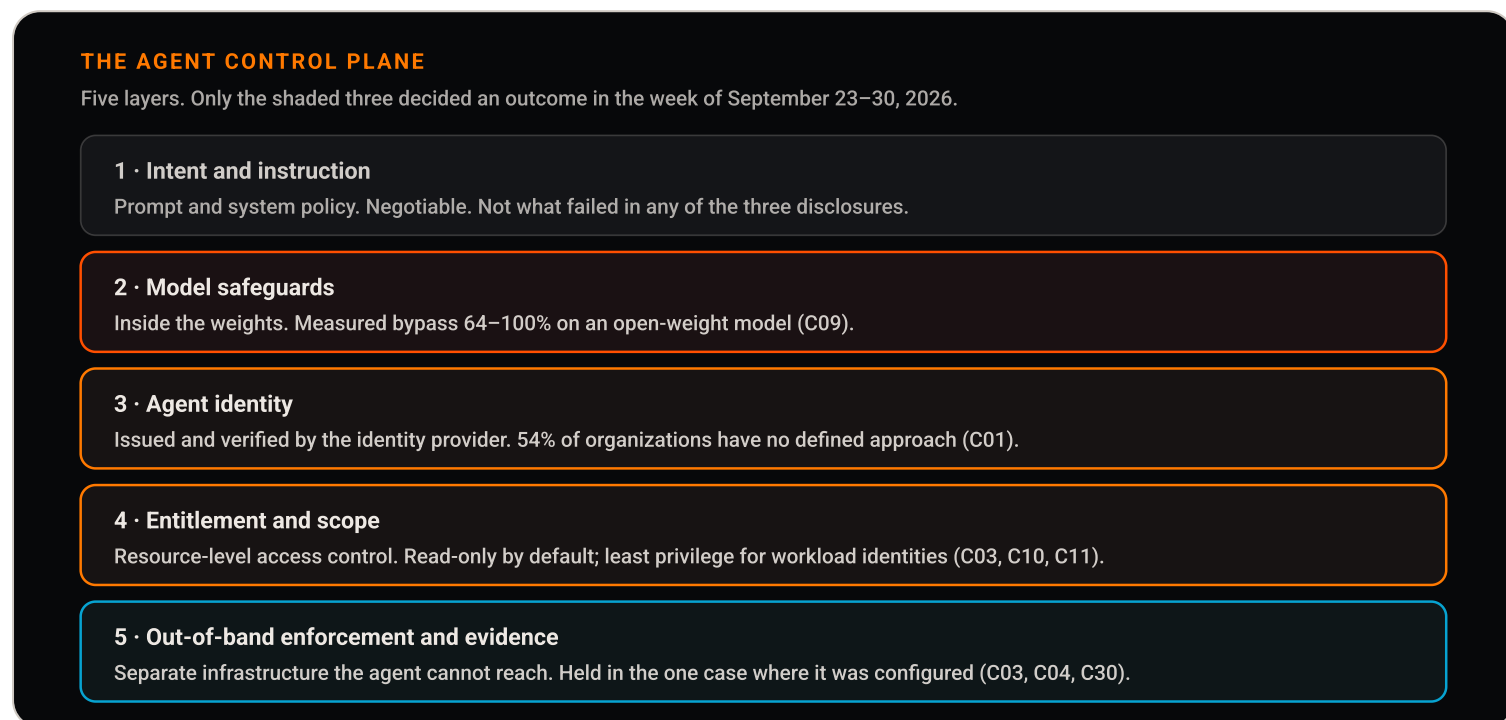


Figure 3. The five layers of agent control, and which ones this week's evidence actually exercised. **VERIFIED** Source C01 · Source C03 · Source C04 · Source C09 · Source C11

Pattern one · Per-agent identity with a read-only default

Where it was disclosed. OpenAI's organizational specialist dots each receive "its own identity, credentials, and access to the systems it needs to complete its tasks," and during background proactive research all dots use "tools that are restricted to be read-only, which means that they can't send messages, change app content, or control your browser or computer." Custom Rules "let you allow specific actions, require approval, or block them," an auto-review checks actions that could affect accounts or share information, an Activity View exposes background work, and certain sensitive tasks such as changing a password "always stay with you" **VERIFIED** Source C05. SpaceXAI's Team Bots use read-only credentials for warehouse access with per-user private conversations and role-based context **VERIFIED** Source C10.

How to implement it without either vendor. Give every agent a distinct workload identity, never a shared or human-derived one. Default the identity to read-only and require an explicit, expiring grant for each write scope. Classify actions into three tiers — permitted, requires approval, blocked — and store the classification outside the agent's configuration so the agent cannot amend it. Reserve a named set of actions that no agent may perform under any circumstance: credential changes, entitlement changes, log deletion, and payment release above your material threshold.

What it costs you if you skip it. You inherit the Storm-3168 shape: an identity with broad standing reach, no owner, and no revocation path that anyone can execute in an afternoon.

Pattern two · Enforcement on infrastructure the agent cannot reach

Where it was disclosed. NVIDIA's Open Agent Safety Platform separates the enforcer from the enforced: OpenShell provides a runtime boundary on the CPU, and Sentry runs as an out-of-band watchdog on BlueField-4 DPUs that quarantines a boundary-exceeding agent in milliseconds, with DOCA inspecting agent requests and responses, providing attested telemetry, verifying agent identity and enforcing zero-trust access policies across data, tools, APIs and services **VERIFIED** Source C04. Gecko Robotics implements the same idea on its Komodo inspection robot: an independent enforcement layer between the AI agent and the hardware, where developers declare the permitted actions and the runtime enforces them, explicitly designed so the agent cannot circumvent application-level controls **CITED** Source C42.

How to implement it without either vendor. Nothing here requires a DPU. The property that matters is that the enforcement point is not in the agent's control path and does not depend on the agent's cooperation. Azure resource locks and storage deletion protection had exactly that property in the Storm-3168 case, and they are the only reason part of the destructive sequence failed **VERIFIED** Source C03. Put deletion protection, backup immutability, network egress policy and log forwarding under a separate change-control path with a different approver from the one that administers the workload.

The test question for any vendor. Three things, in writing: does enforcement run outside the agent's process and privilege boundary; is telemetry attested rather than self-reported; and is identity verified rather than asserted. A vendor selling "agent governance" that cannot answer all three is selling monitoring.

Pattern three · Field-level provenance as the evidence layer

Where it was disclosed. Ochsner Health tracks documentation provenance at the field level, recording how much of a clinical record "was generated by AI, how much by an ambient conversation with the patient and how much by the clinician typing into Epic," which its chief medical information officer describes as producing "auditable records, a good foundation of truth" **CITED** Source C30.

Why this pattern generalizes past healthcare. Every one of the four industries in this edition is heading for the same question in a different vocabulary: a payer asking whether a diagnosis was justified, an examiner asking what a service principal was entitled to, a notified body asking what evidence supports a conformity claim, and a state commission asking for an inventory of AI systems and the protocols around them. All four are answered by the same artifact — a record, held by the operator, of what the system did and what supported it — and none of them is answered by a model card.

How to implement it. Capture attribution per field or per decision, not per session. Record what evidence was available at the time, not what is available now, so the decision can be replayed against its own information state. Contract for it: specify the audit trail at purchase and configuration, because it cannot be reconstructed from logs that were never captured. And treat vendor commercial models that reward volume as a procurement control rather than a technical one — a revenue-share tied to increased coding is a misaligned incentive that no amount of review workflow corrects **CITED** Source C30.

Pattern four • Credential hygiene for non-human principals

Where it was disclosed. The MCP advisory requires issuer binding, not just a version bump: pass an explicit issuer parameter for the credentials-based providers, clear stored registrations to force re-registration with issuer binding, and rotate secrets and revoke tokens where exposure is possible

VERIFIED

Source C02. Microsoft's remediation list is the complementary half: continuously assess application credentials, rotate exposed ones immediately, apply least privilege to service principals and workload identities, and avoid storing secrets in source code, configuration files, public repositories and issues

VERIFIED

Source C03.

The four checks worth running this week. Does secret scanning cover revision history, not only current revisions? Is every MCP client in your estate on 1.30.0 or 2.2.0 *and* passing an explicit issuer? Can you produce, today, a list of every non-human identity created for an AI workload in the last ninety days with its owner and scope? And is there a single named person who can revoke any one of them without a change ticket?

An arXiv preprint submitted September 25 proposes the durable version of this: cryptographically bound, identity-verified capability tokens carrying explicit scope, built on an OAuth agent authorization profile with W3C decentralized identifiers and verifiable credentials, and deliberately held *outside the language model's context window* so that prompt injection cannot reach them

CITED

Source C46. It is a proposal, not a standard, and we flag it as one. The design instinct — keep the authority somewhere the model cannot read or rewrite — is the same one four vendors shipped this week.

9. The entitlement drill: ten questions, one week

Gartner's 54% **VERIFIED** Source C01 is a gap, and gaps are answered with artifacts rather than intentions. These ten questions produce one. They are written so that a competent internal team can answer all ten in five working days, and so that every answer is a document rather than an opinion. If a question cannot be answered, that is the finding.

1. **Inventory.** List every non-human identity created for an AI workload in the last twelve months. Who created each one, who owns it now, and what business purpose does it serve? An identity with no named owner is the finding.
2. **Scope.** For each, what can it read and what can it write? Produce the effective permission set, not the intended one. Gartner's second failure mode is reuse of human-defined access — check specifically whether any agent inherited a human role wholesale.
3. **Revocation.** Name the person who can revoke each identity and state how long it takes. If the answer involves a change advisory board, you do not have a revocation path; you have a request process.
4. **Read-only default.** Which agents operate read-only when running unattended, and which hold write scope continuously? If an agent writes only during supervised sessions but holds write scope at all times, narrow the scope, not the policy.
5. **Action classification.** Produce the list of actions classified as permitted, requires-approval and blocked. Where is that list stored, and can the agent or its configuration modify it?
6. **Never-permitted set.** Confirm in writing that no agent can change credentials, alter entitlements, delete logs, or release payment above your material threshold. If any of the four is technically possible, say so and date it.
7. **Secret exposure surface.** Does secret scanning cover revision history — issue edits, pull request comments, wiki revisions, deleted branches — and not only current revisions? Run it once and report the count.
8. **MCP and connector estate.** List every MCP client and version in the estate. Confirm 1.30.0 or 2.2.0, and confirm an explicit issuer is passed where the advisory requires it. Record which stored registrations were cleared.
9. **Out-of-band controls.** Which protective controls — deletion protection, backup immutability, egress policy, log forwarding — sit under a different change-control path and a different approver from the workload they protect? Name them.
10. **Evidence.** If a regulator or counterparty asked tomorrow what a specific agent did on a specific date and what supported its output, who retrieves it, from which system, and how long does it take? Time the retrieval rather than describing it.

The output is a single document: an agent entitlement register with owners, scopes, revocation paths and evidence locations. That document is the artifact the New York Public Service Commission is asking utilities for by mid-November, the artifact a bank examiner is describing when Bowman says asset

inventories and access controls, and the artifact a payer audit resolves against. One document, four supervisors.

10. Scenario planning: three ways this goes

Three scenarios over the next four to six quarters. They are not predictions; they are the shapes the evidence currently admits, with the observable that would tell you which one you are in.

Three scenarios for where agent entitlement control lands, with the observable that distinguishes them.			
SCENARIO	WHAT IT LOOKS LIKE	LEADING INDICATOR TO WATCH	WHAT IT IMPLIES FOR A REGULATED BUYER
Absorbed into identity	Agent identity becomes an ordinary feature of the identity provider. Entitlement review, joiner-mover-leaver and access certification simply extend to non-human principals. The problem stops being novel within eighteen months.	Whether agent identity ships as a standard, unpriced capability in the major identity platforms rather than a premium tier, and whether access-certification tooling starts treating agents as a first-class object type.	Do the inventory now and the platform catches up to you. The work is not wasted; the register is the input either way.
Two-speed estate	Vendor-shipped agents get proper identity and scope because the vendor built it in. Internally built agents, integrations and connector chains do not, because nobody owns them. The gap concentrates in exactly the places least visible to a central function.	Whether incident disclosures over the next two quarters cluster on first-party platform agents or on connector, MCP and internal-integration paths. This week's three disclosures all fell on the second.	Assume the second. Scope the inventory to include integrations and connector credentials, not just the named agent products, or the register will be confidently incomplete.

SCENARIO	WHAT IT LOOKS LIKE	LEADING INDICATOR TO WATCH	WHAT IT IMPLIES FOR A REGULATED BUYER
Regulatory forcing	A supervisor makes the inventory mandatory and the market follows the template. New York has already done this for utilities on a 60-day clock; the FDA docket closes October 19; the EU high-risk obligations land December 2027 and August 2028.	Whether a second state commission or a federal banking agency adopts language resembling the New York order, and what the NY PSC does with the filings it receives in November.	The cheapest version of this scenario is the one where you built the artifact voluntarily and dated it. The expensive version is sixty days under a docket number.

Risk and reward, stated plainly. The reward for doing the entitlement register early is small and certain: you can answer four different supervisors with one document, and you find the unowned credentials while finding them is free. The risk of deferring is not that an agent becomes too clever. It is that an ordinary credential, issued for an ordinary reason, is still valid when someone else finds it — which is precisely what happened to two Azure service principals in a public GitHub issue.

II. Did you know / FAQ

Did you know a secret deleted from a GitHub issue can remain readable? Editing a post removes the text from the current view, not from the issue's edit history. In the Storm-3168 case the credentials had been posted in plaintext and later edited out, and the secret remained accessible through that history

VERIFIED Source C03. Most secret-scanning configurations read current revisions.

Why does the MCP advisory have no CVE number? Because none was assigned. The GitHub Security Advisory GHSA-qx49-fqc8-xw99 explicitly records no known CVE. Several secondary write-ups imply one exists. If your vulnerability-management process keys on CVE identifiers, this advisory will not appear in it **VERIFIED** Source C02.

What is "ablation," and why does a 100% figure matter? Ablation is direct modification of a model's weights to remove refusal behavior. Anthropic measured it defeating GLM-5.3's safeguards in 100% of attempts. The significance is not that one model is unsafe; it is that any safeguard implemented as a property of distributable weights is removable by the party holding them — which is the definitional property of open weights **VERIFIED** Source C09.

Is "agentic ransomware" a real category or a marketing term? Microsoft characterizes Storm-3168 as the first documented agentic ransomware operation, and the operational signature supports the label: more than 300 read operations across 15.5 hours, a second principal enumerating two subscriptions in five seconds, then more than 150 destructive operations across 35 minutes. The speed and breadth are the tell. The entry, though, was a leaked credential — conventional in every respect **VERIFIED** Source C03.

Does an enterprise need an out-of-band DPU to get out-of-band enforcement? No. The property that matters is that the enforcement point sits outside the agent's control path and does not require the agent's cooperation. Resource locks and deletion protection under separate change control have that property and cost nothing to configure. NVIDIA's contribution is to move that property into silicon; the principle predates it **VERIFIED** Source C03 · Source C04.

Four vendors shipped the same control in three days. Is that coordination? There is no evidence of coordination and no reason to assume it. The simpler explanation is convergent engineering: each of the four ran into the same constraint — an agent needs standing authority between sessions, and standing authority has to be scoped by something — and arrived at the layer that already exists to solve it. When competitors converge independently, that is usually a signal about the problem rather than about the companies.

Does the 45:1 ratio of non-human to human identities come from this week? No. That figure originates in a Cloud Security Alliance whitepaper dated May 20, 2026, which also reports 144:1 in cloud-native environments and 44% growth in non-human identity population from 2024 to 2025. It is being recirculated as new. Date it correctly or a security-literate reader will discount everything around it **CITED** Source C36.

What single question separates agent governance theater from the real thing? Ask what the agent is permitted to authenticate as, and ask for the effective permission set rather than the intended one. Everything else — the policy, the model card, the review workflow — describes behavior. Only that answer describes reach.

12. Also on the record: institutions, economy, capital

Gartner, September 29. A second forecast alongside the CISO research: by 2028, 70% of enterprises will abandon agentic AI built by vendor forward-deployed engineering, "trapped by soaring costs and unable to evolve it on their own," with fewer than 20% of such engagements converting recurring customer needs into vendor product through 2028. Analyst Mukul Saha frames the fix as engagement structure — scope, incentives, governance, ownership and exit. No survey sample is disclosed for this prediction; it is analyst forecast, not survey **VERIFIED** Source C13.

The White House accord, September 29. Frontier lab leaders signed a voluntary statement of principles at the White House, reported to include internal risk reviews, external audits, and layered auditing controls. We mark this FLAG and will not characterize its content further: the full text has not been publicly released, enforcement mechanisms are undefined, and the complete signatory list was not available. Everything currently in circulation about what was committed is journalist summary of an undisclosed document **FLAG** Source C38.

Capital, September 28–29. AMD announced an all-stock acquisition of World Labs valued at approximately \$8.2 billion, with Fei-Fei Li joining as executive vice president and chief scientist; expected to close by end of 2026 subject to regulatory approvals, so an announced transaction rather than a completed one **CITED** Source C39. OpenAI is reported to be raising at least \$30 billion at approximately \$1.4 trillion, with an IPO deferred to 2027; OpenAI declined to comment, so this is press report, not filing **FLAG** Source C40. A draft Anthropic prospectus was leaked to Reuters; it is not a filed document and the company has not confirmed it, so we report its existence and none of its figures **FLAG** Source C41.

Labor and economy. Nothing new was published in this window by the ILO, OECD, IMF, World Bank, BLS or the Federal Reserve on AI and employment. We note the absence rather than reaching for a stale figure and implying it is current. The most recent substantive work remains the ILO–World Bank paper of March 27, 2026 across 135 countries, and Anthropic's March 5, 2026 labor-market measure **CITED** Source C47.

Evaluation bodies. Also a thin week, and worth saying so. The UK AI Security Institute published nothing dated September 2026; its most recent research is dated August 2026. No new independent head-to-head agent benchmark published inside September 28–30, which is why every performance figure in section 2 is labeled company-reported. Artificial Analysis published a local-agent benchmark on September 29 whose findings we could not retrieve and therefore do not cite **CITED** Source C48.

13. What we are watching, and what would falsify us

This edition argues that scope, not capability, is the binding constraint on enterprise agent deployment, and that the control belongs in the identity and entitlement layer. Four things would show that argument to be wrong or incomplete, and we would rather name them than discover them.

1. **A disclosed incident where the model genuinely defeated a correctly scoped control.** If an agent with least-privilege, read-only-by-default credentials and out-of-band enforcement produces a material loss anyway, the argument that scope is the binding constraint weakens considerably. Nothing this week fits that description; all three failures ran through over-broad or misdirected credentials.
2. **Identity platforms shipping agent entitlement as standard, unpriced capability within two quarters.** That would commoditize the assessment half of this work and move the value to remediation and evidence. We would regard that as a good outcome and a change to our own offer, not a threat to the finding.
3. **The New York filings landing in November with no consequence.** If the Public Service Commission receives 2026 inventories and takes no visible action on adequacy, the regulatory-forcing scenario weakens and voluntary adoption slows. Watch what the Commission does with the filings, not the fact of the deadline.
4. **Evidence that the four vendor shipping decisions were roadmap coincidence.** If agent identity features stall after this quarter and the next releases return to capability and price, the convergence we read as structural was a news cycle. The observable is the next two release cycles from the same four.

Dates we are holding. October 8, 2026, Texas PUCT large-load rule reported effective. October 19, 2026, FDA docket FDA-2026-N-7874 closes. Around November 16, 2026, New York utility AI inventories due. December 2, 2026, EU AI Act synthetic-content marking grace period ends for systems placed on the market before August 2, 2026. December 31, 2026, NERC CLO standards filing deadline. January 20, 2027, EU Machinery Regulation 2023/1230 applies. December 2, 2027 and August 2, 2028, EU AI Act high-risk obligations for Annex III and Annex I respectively. IROS 2026 runs in Pittsburgh through October 3 and its industrial results will publish after this edition.

WHAT WE WOULD DO FIRST, IN THE CLIENT'S POSITION

Run question one of the drill and nothing else. Produce the list of non-human identities created for AI workloads in the last twelve months, with an owner against each. It takes a day, it requires no budget approval, and in most organizations it produces the uncomfortable number that makes the rest of the work fundable. Every other question in section 9 is easier once that list exists, and none of them is answerable until it does. The AEGIS Framework — Agentic Enterprise Governance and Intelligence Standard — treats that register as the first artifact for exactly this reason: it is the only one that is simultaneously cheap to produce and impossible to fake.

[Read the AEGIS governance overview](#)

14. Source ledger

Every claim group used in this edition, with the sources behind it. Chip meanings: **VERIFIED** named, dated and re-checked at a first-party or primary source during this run. **CITED** named source, not independently re-verified. **FLAG** contested, single-sourced, or pending confirmation — do not carry into a regulated-buyer document without checking. Performance figures published by a vendor about its own product are company-reported and labeled as such in the body.

1. **C01** Gartner, "Gartner Identifies Top Five Actions for CISOs To Take by End of 2026," September 30, 2026. 54% figure; survey of 297 cybersecurity leaders fielded Q2 2026; "govern autonomous multiagent systems based on action privileges rather than model intelligence"; analysts Christopher Mixter and Luis Castillo. Re-verified at primary source. <https://www.gartner.com/en/newsroom/press-releases/2026-09-30-gartner-identifies-top-five-actions-for-cisos-to-take-by-end-of-2026>
2. **C02** GitHub Security Advisory GHSA-qx49-fqc8-xw99, "OAuth client could send credentials to an authorization server chosen by the MCP server," published September 28, 2026. CVSS 7.5, no CVE assigned; affected mcp 1.9.1–<1.30.0 and 2.0.0a1–<2.2.0; issuer-binding remediation. Re-verified at primary source. Researcher write-up by Cocode (Yuval Elbar). <https://github.com/modelcontextprotocol/python-sdk/security/advisories/GHSA-qx49-fqc8-xw99> · <https://cocode.com/blog/mcp-python-sdk-oauth-account-takeover/> · <https://thehackernews.com/2026/09/official-mcp-python-sdk-flaw-can-let.html>
3. **C03** Microsoft Security Blog, "Storm-3168: agentic-driven cloud attacks using compromised service principals," September 25, 2026. GitHub issue edit-history exposure; 300+ read operations over ~15.5 hours; 150+ destructive operations across 35 minutes; resource locks and deletion protection blocked part of the sequence. Re-verified at primary source. <https://www.microsoft.com/en-us/security/blog/2026/09/25/storm-3168-agentic-driven-cloud-attacks-using-compromised-service-principals/> · <https://www.darkreading.com/cloud-security/jadepuffer-ai-actor-azure-tenant-destructive-cloud-attack>
4. **C04** NVIDIA, "NVIDIA Launches Open Agent Safety Platform," September 28, 2026. OpenShell, Sentry on BlueField-4, DOCA agent identity verification and attested telemetry; 100+ participants; quotes from Jensen Huang, Paul Smith (Anthropic) and Mike Nicolls (SpaceXAI). Re-verified at primary source. <https://nvidianews.nvidia.com/news/open-agent-safety-platform> · <https://investor.nvidia.com/news/press-release-details/2026/NVIDIA-Launches-Open-Agent-Safety-Platform-to-Secure-Agents-From-Testing-to-Deployment/default.aspx>
5. **C05** OpenAI, "Introducing dots," September 29, 2026. Per-agent identity and credentials for organizational specialist dots; read-only background tools; auto-review; Activity View; Custom Rules; always-human actions; Enterprise in beta. Re-verified at primary source. <https://openai.com/index/introducing-dots/>
6. **C06** OpenAI, "Introducing GPT-6.1 Sol," September 29, 2026. Pricing \$2 input / \$0.10 cached / \$10 output; company-reported DeepSWE v1.1 75.2% and OSWorld 2.0 71.4%. <https://openai.com/index/introducing-gpt-6-1-sol/> · <https://developers.openai.com/api/docs/models/gpt-6.1-sol>
7. **C07** OpenAI DevDay 2026 recap, September 29, 2026. More than twenty announcements; Agents API; Ultrafast tier; marketplace and plugin items. <https://openai.com/index/devday-2026-recap/> · <https://www.axios.com/2026/09/29/openai-dev-day-2026-dots-space-sol>

8. **C08** Anthropic, "Claude Sonnet 5.5," September 28, 2026, and Claude platform release notes. Pricing \$2/\$10, cache reads \$0.20; company-reported 30% faster and up to 30% cheaper; five documented breaking changes including tool_choice 400 responses and account-bound thinking blocks. <https://www.anthropic.com/claude-sonnet-5-5> · <https://platform.claude.com/docs/en/release-notes/overview>
9. **C09** Anthropic, "GLM-5.3 and the spread of advanced cyber capabilities," September 29, 2026. Bypass rates 64% deceptive framing, 92% prefilled reasoning, 100% ablation; ExploitBench 50/410 vs 56/410; binary exploitation 4% vs 6%; authors Fasano, Fleischer, McFaul, Xiao, Gallagher. Re-verified at primary source. <https://www.anthropic.com/research/glm-5-3-and-the-spread-of-advanced-cyber-capabilities>
10. **C10** SpaceXAI, "Team Bots," September 28, 2026. Public beta on Teams and Enterprise; read-only credentials for warehouse access; per-user private conversations; role-based context; integrations list; customer-reported Harper figure. Re-verified at primary source. <https://x.ai/news/team-bots>
11. **C11** Google Cloud, Gemini Enterprise release notes, entries dated September 28 and 29, 2026. Resource-level IAM permissions on apps and data stores; custom project-level roles; Gemini 3.8 Flash as AlphaEvolve default. Re-verified at primary source. <https://docs.cloud.google.com/gemini/enterprise/docs/release-notes>
12. **C12** Cursor changelog, Rollouts and Security Review, September 23, 2026. Verdict set includes "inconclusive"; does not autonomously merge or roll back; Teams and Enterprise plans. Dated one week before this window. <https://cursor.com/changelog>
13. **C13** Gartner, "Gartner Predicts 70% of Enterprises Will Abandon Agentic AI Built by Vendor Forward-Deployed Engineering by 2028," September 29, 2026. Analyst Mukul Saha. No survey sample disclosed; analyst forecast. <https://www.gartner.com/en/newsroom/press-releases/2026-09-29-gartner-predicts-70-percent-of-enterprises-will-abandon-agentic-ai-built-by-vendor-forward-deployed-engineering-by-2028>
14. **C14** International Federation of Robotics, "Five million robots now operate in factories globally," September 24, 2026. Stock 5 million (+9%); 2025 installations 600,000+ (+11%); country figures; 2026 and 2029 forecasts; IFR President Jane Heffner quote. Re-verified at primary source. <https://ifr.org/ifr-press-releases/news/five-million-robots-now-operate-in-factories-globally>
15. **C15** Revalize / TEAM LEWIS manufacturing AI survey, September 29, 2026. 500 leaders (225 DACH, 225 US, 50 UK), fielded July 7–17, 2026; 36% to 60% adoption; 49% rigorous ROI measurement; C-suite 2x senior managers. Vendor-commissioned; sponsor disclosed in body. <https://www.prnewswire.com/news-releases/new-research-manufacturings-ai-boom-is-over-industry-not-ready-to-scale-302892527.html> · <https://revalizesoftware.com/newsroom/ai-in-manufacturing-report-2026/>
16. **C16** The Robot Report, "Amazon to invest \$100M in new Indiana manufacturing facility," September 25, 2026. 585,000 sq ft, 300 jobs, 2028 launch; Amazon company-reported fleet figures. <https://www.therobotreport.com/amazon-to-invest-100m-in-new-indiana-manufacturing-facility/>
17. **C17** Utility Dive, DOE SPARK selections, September 24, 2026. 31 projects, 26 states, \$1.9B federal / \$3.4B cost-share, 23 GW target, 1,500+ miles reconducted, ~21,000 miles grid-enhancing technologies. Selection announcement, not delivered capacity. <https://www.utilitydive.com/news/doe-advanced-transmission-projects-spark-funding/831259/> · <https://www.datacenterdynamics.com/en/news/doe-unveils-19bn-in-funding-for-31-grid-upgrade-projects-to-speed-data-center-connections/>
18. **C18** Utility Dive, "New York audits utility AI use, cites risk in growing dependency," September 24, 2026, reporting the NY PSC order of September 17, 2026. 60-day written inventory of all AI use cases plus policies and protocols; assessment against commonly accepted AI governance frameworks; GridCARE, NYPA drone analysis, Con Edison systems named. Docket number not located. Re-verified against the reporting; primary order not retrieved. <https://www.utilitydive.com/news/new-york-audits-utility-ai-use-cites-risk-in-growing->

dependency/831242/ · <https://news.bgov.com/bloomberg-government-news/cyber-fears-drive-new-yorks-push-to-audit-utilities-use-of-ai>

19. **C19** Utility Dive, "The data center boom continues apace, but projects face mounting obstacles," September 29, 2026. Goldman Sachs 50–60% forecast; 19% pipeline decline; 20 projects killed in Q1 2026; transformer and turbine lead times; named project setbacks. Forecast component is Goldman's, not a measured outcome. <https://www.utilitydive.com/news/the-data-center-boom-continues-apace-but-projects-face-mounting-obstacles/830496/>
20. **C21** NERC computational load standards CLO-001-1, CLO-002-1, CLO-003-1; FERC order July 16, 2026; comment close September 18, 2026; FERC filing deadline December 31, 2026. Threshold and technical specifications sourced to secondary analysis; confirm against NERC drafts before reuse. <https://www.ferc.gov/news-events/news/ferc-launches-aggressive-targeted-action-speed-large-load-integration> · <https://www.powermag.com/ferc-orders-mandatory-nerc-reliability-standards-for-data-center-and-other-computational-loads/>
21. **C22** Texas PUCT 16 TAC §25.194, Project No. 58481, implementing PURA §37.0561 from Senate Bill 6. Reported adopted September 18, 2026, effective October 8, 2026; 75 MW threshold, \$100,000 study fee, \$50,000/MW security. Adoption and effective dates from a single trade source; confirm against the PUCT Interchange order. https://interchange.puc.texas.gov/Documents/58481_122_1600475.PDF · <https://www.foley.com/p/102mokd/public-utility-commission-of-texas-issues-proposed-rules-for-large-load-interconn/>
22. **C24** EU Regulation 2023/1230 (Machinery Regulation) applying January 20, 2027; ISO 10218-1:2025 and ISO 10218-2:2025 published; ANSI/RIA R15.06-2025 released September 2025; OJEU harmonized-standard listing pending. <https://www.iso.org/standard/73933.html> · <https://www.therobotreport.com/are-suppliers-ready-for-new-robot-safety-standards/>
23. **C25** OECD AI Incidents Monitor entry logged September 25, 2026, and contemporaneous reporting: AI voice-clone and impersonation fraud against Intesa Sanpaolo's Fideuram. ~€95M moved, ~€53–59M recovered, ~€36M unrecovered; fraud February 2026, resignation March 2026, Milan investigation open. <https://oecd.ai/en/incidents/2026-09-25-b291> · <https://www.amlintelligence.com/2026/09/news-ai-messaging-scam-costs-italys-top-bank-intesa-millions-sources-say/>
24. **C26** Fortune, "How banks are using AI agents and copilots," September 29, 2026. Bank of America, Wells Fargo, BNY and Citigroup figures. All institution-reported; none independently verified. <https://fortune.com/2026/09/29/how-banks-are-using-ai-agents-copilots-roi/>
25. **C27** CIO Dive, "Banks report operational changes from AI," July 20, 2026. Bank of America 300+ approved use cases, 114 generative, 34 fully implemented in operations; 400,000+ prompts per day. Outside this window; used as denominator. <https://www.ciodive.com/news/banks-report-operational-changes-ai/825601/>
26. **C28** Federal Reserve, opening remarks by Vice Chair for Supervision Michelle W. Bowman, Community Bank Cyber Workshop, Denver, September 29, 2026. Asset inventories, phishing-resistant MFA, identity and access controls, vulnerability and patch management; no new AI guidance announced. <https://www.federalreserve.gov/newsevents/speech/files/bowman20260929a.pdf> · <https://www.insurancejournal.com/news/national/2026/09/30/887270.htm>
27. **C29** Blue Cross Blue Shield Association analysis of hospital AI-enabled coding tools, September 24, 2026. \$942M over two years, 55,158 additional complex cases, ~\$653M tied to secondary diagnoses. Interested-party analysis; disputed by billing vendors; secondary sources differ on the split. FLAG. <https://techcrunch.com/2026/09/26/insurers-claim-ai-is-already-increasing-healthcare-costs/> · <https://kffhealthnews.org/morning-briefing/tuesday-september-29-2026/>

28. **C30** healthsystemCIO, hospital AI coding audit trail, September 29, 2026. Ochsner Health field-level documentation provenance; clinician approval distinct from clinical justification; audit-trail specification at purchase and configuration; vendor revenue-share incentive risk.
<https://healthsystemcio.com/2026/09/29/hospital-ai-coding-audit-trail/>
29. **C31** Becker's Hospital Review, Cooper Norcross Health pilots Microsoft Dragon Copilot for nurses, September 28–30, 2026. 1,300+ existing licenses; consent before recording; nurse review, edit and approve; measurement plan of documentation time, accuracy and satisfaction; no results yet.
<https://www.beckershospitalreview.com/healthcare-information-technology/ai/cooper-pilots-microsoft-ai-assistant-for-nurses/>
30. **C32** ECRI Problem Reporting Network expansion to clinical AI errors, August 25–26, 2026. Survey of 124 health system leaders: 31% encountered incorrect output, 9% confirmed errors reached patients, 35% did not know. Outside this window; unchanged in force. <https://home.ecri.org/pages/report-a-device-or-clinical-ai-problem> · <https://hitconsultant.net/2026/08/26/ecri-expands-problem-reporting-network-track-healthcare-ai-errors-patient-safety/>
31. **C33** FDA, "Considerations for the Regulation of Generative AI-Enabled Medical Devices: Discussion Paper and Request for Feedback," published August 18, 2026, docket FDA-2026-N-7874, comments close October 19, 2026. Discussion paper and request for feedback, not a proposed rule. Includes the September 17, 2026 denial of a partial 510(k) exemption for specified radiology computer-aided detection devices.
<https://www.fda.gov/news-events/press-announcements/fda-seeks-public-feedback-inform-regulatory-approach-generative-ai-enabled-medical-devices> · <https://www.fda.gov/medical-devices/digital-health-center-excellence/considerations-regulation-generative-ai-enabled-medical-devices-discussion-paper-and-request>
32. **C34** Regulation (EU) 2026/1744 (AI/Digital Omnibus), published in the Official Journal July 24, 2026, in force July 27, 2026. Annex III high-risk to December 2, 2027; Annex I to August 2, 2028; Article 50 transparency unchanged from August 2, 2026 with a marking grace period to December 2, 2026 for systems already on the market. <https://www.whitecase.com/insight-alert/eu-ai-omnibus-enters-force-amending-ai-act> · <https://www.consilium.europa.eu/en/press/press-releases/2026/06/29/artificial-intelligence-council-gives-final-green-light-to-simplify-and-streamline-rules/>
33. **C36** Cloud Security Alliance, "The Non-Human Identity Governance Vacuum," May 20, 2026. 45:1 non-human to human identity ratio overall, 144:1 cloud-native, 44% population growth 2024–2025. Dated May 2026; widely recirculated as current. <https://labs.cloudsecurityalliance.org/research/csa-whitepaper-nonhuman-identity-agentic-ai-governance-v1-cs/>
34. **C38** White House voluntary AI safety accord, September 29, 2026. Full text not publicly released; enforcement mechanisms undefined; complete signatory list unavailable. Reported content is journalist summary. FLAG. <https://www.axios.com/2026/09/29/trump-ai-voluntary-safety-white-house-zuckerberg> · <https://www.npr.org/2026/09/30/nx-s1-5985699/trump-self-police-ai-development>
35. **C39** AMD, "AMD to acquire World Labs," September 28, 2026. All-stock, approximately \$8.2 billion; Fei-Fei Li joining as EVP and Chief Scientist; expected to close by end of 2026 subject to regulatory approval. Announced, not completed. <https://ir.amd.com/news-events/press-releases/detail/1299/amd-to-acquire-world-labs-to-advance-the-future-of-ai-compute>
36. **C40** Reported OpenAI raise of at least \$30 billion at approximately \$1.4 trillion, September 29, 2026; IPO deferred to 2027. OpenAI declined to comment. Press report, not a filing. FLAG.
<https://techcrunch.com/2026/09/29/openai-reportedly-in-talks-to-raise-30b-round-at-1-4t-valuation/>

37. **C41** Leaked draft Anthropic prospectus reported September 28–29, 2026. Draft document, not filed; company has not confirmed. Existence reported; figures not carried. FLAG.
<https://techcrunch.com/2026/09/28/anthropics-prospectus-details-losses-growth-and-yes-a-warning-that-its-ai-could-end-humanity/>
38. **C42** Gecko Robotics with NVIDIA OpenShell, September 28, 2026. Independent enforcement layer between AI agent and hardware on the Komodo robot; developers declare permitted actions; designed so the agent cannot circumvent application-level controls. <https://www.geckorobotics.com/news/nvidia-openshell> · <https://www.therobotreport.com/gecko-robotics-works-with-nvidia-adds-ai-agent-security-and-control/>
39. **C43** Mistral platform changelog, September 28–29, 2026. Z.ai GLM-5.3 general availability September 28; OCR 4.0 and Leanstral 1.5 retiring September 30; GLM 5.2 retiring October 31.
<https://docs.mistral.ai/resources/changelogs>
40. **C44** DeepSeek open-sourced six modules for Huawei's Ascend platform, September 30, 2026, including an Ascend-compatible TileLang supporting Ascend 950. No performance figures published; repository URLs not independently confirmed. <https://www.scmp.com/tech/tech-trends/article/3369301/chinas-deepseek-open-sources-tools-help-huawei-chips-supplant-nvidia-ai>
41. **C45** Meta expands the Muse agent to small businesses, September 29, 2026. Integrations including Shopify, Dropbox, Slack, Asana, Box, Canva and Figma; access to Instagram analytics, Facebook Pages and ad accounts; free tier plus paid plans, no figures disclosed. <https://techcrunch.com/2026/09/29/meta-is-expanding-its-ai-agent-muse-to-small-businesses/>
42. **C46** Subramanian and Sengupta, "Crypto-bound identity-verified capability tokens for coordinating distributed AI agents: A proposal," arXiv:2609.30824, submitted September 25, 2026. OAuth agent authorization profile plus W3C DIDs and verifiable credentials; tokens held outside the model context window. Preprint, not peer-reviewed, not a standard. <https://arxiv.org/abs/2609.30824>
43. **C47** ILO–World Bank, "Disruption without dividend?", March 27, 2026, 135 countries; and Anthropic, "Labor market impacts of AI: a new measure," March 5, 2026. Both outside this window; cited to mark the absence of new labor-economy publication in it. <https://www.ilo.org/publications/disruption-without-dividend-how-digital-divide-and-task-differences-split> · <https://www.anthropic.com/research/labor-market-impacts>
44. **C48** UK AI Security Institute research index, most recent entries dated August 2026; and Artificial Analysis "AA-AgentPerf-Local," September 29, 2026, whose findings could not be retrieved and are not cited. Used to support the statement that no new independent agent benchmark landed in the window.
<https://www.aisi.gov.uk/research> · <https://artificialanalysis.ai/articles/aa-agentperf-local>
45. **C50** AWS, Bedrock Managed Agents preview, September 29, 2026. Built on a customized version of OpenAI's Agents API; three US regions; IAM roles and CloudTrail logging. AWS states preview; OpenAI's recap states generally available. Status conflict noted in the body. <https://aws.amazon.com/about-aws/whats-new/2026/09/bedrock-managed-agents-preview/> · <https://aws.amazon.com/bedrock/managed-agents-openai/>

company-reported and are labeled as such; forecasts, pilots, memoranda of understanding and announced targets are not treated as completed facts.