SKYBREAK · WHITE PAPER · SEPTEMBER 2026

The Ungoverned Plug

MCP connected AI to everything — and governed none of it. The gaps in today's MCP market, and how SkyBreak's Governed MCP closes them

In under two years, the Model Context Protocol became the universal way to plug AI into enterprise tools — ten thousand public servers, ninety-seven million monthly SDK downloads, first-party support from every major AI vendor. It also became the fastest-growing attack surface in enterprise software: measured studies find command-injection flaws in 43% of servers, static API keys in more than half, OAuth adoption in single digits, and thousands of servers with at least one security finding. This paper surveys the MCP market as it stands, documents the gaps with published evidence, and presents SkyBreak's answer: the NEXUS Governed MCP — the same universal plug, behind an enterprise breaker panel of admission contracts, bounded authority, independent verification, and audit-grade receipts, deployed privately with enterprise-specific SLMs.

Autonomy you can prove.

Prepared by SkyBreak · skybreak.us · Confidential — for discussion

Executive summary

The Model Context Protocol solved a real problem brilliantly: instead of building a custom integration between every AI application and every tool, the industry now has one open standard — introduced by Anthropic in late 2024, donated to the Linux Foundation's Agentic AI Foundation in December 2025, and adopted natively by Claude, ChatGPT, Gemini, Microsoft Copilot, and the major developer platforms. The ecosystem is enormous and growing: roughly 10,000 active public servers, nearly 16,000 open-source server repositories, and an official registry approaching ten thousand entries. Enterprises are following — a 2026 survey found 41% already running MCP in production.

But MCP standardized the plug, not the power system behind it. The protocol answers 'how does an AI reach a tool?' and deliberately leaves open who may use the tool, for what, with whose data, under whose watch, and with what proof. The market's own roadmap admits it: standardized audit trails, multi-tenancy isolation, and gateway authorization behavior are open items, not shipped features. Meanwhile the measured security record is stark — 43% of tested servers carry command-injection flaws, 82% of implementations use file operations prone to path traversal, only 8.5% of credentialed servers use OAuth, and researchers found over a thousand servers listening on the public internet with no authentication at all. A first generation of MCP gateways and registries is emerging to help, but they largely manage connections; none makes an agent's actions provable.

SkyBreak's position: the enterprise doesn't need a different plug — it needs governance behind the plug. The NEXUS Governed MCP speaks the open standard on both sides, so every MCP-compatible client and server still works. What changes is everything in between: servers are admitted by contract with provenance and scope, not installed from a registry on faith; every tool call passes a live, bounded, revocable Authority Grant instead of riding a static API key; results are independently verified against postconditions instead of trusted on the server's word; every call produces an audit-grade receipt with replay; a human holds an emergency stop; and the whole gateway deploys inside the client's own boundary alongside enterprise-specific private SLMs. MCP made AI connected. NEXUS makes it accountable.

The one-line POV: MCP is the USB-C of AI — and the enterprise is plugging thousand-watt agents into it with no breaker panel. SkyBreak sells the breaker panel.

1 · The MCP market today: everyone standardized the plug

The ecosystem has consolidated with remarkable speed. On the demand side, every major AI client is MCP-native: Claude, ChatGPT, Gemini, Microsoft Copilot, GitHub Copilot, Cursor, Windsurf, VS Code, and Zed. On the supply side, first-party servers ship from Slack, GitHub, Google, Salesforce, Stripe, HubSpot, Shopify, Notion, Linear, Sentry, Figma, and Cloudflare, atop a long tail of roughly 10,000 active public servers and ~16,000 open-source server repositories. SDK downloads run at 97 million per month. Governance of the spec itself moved to neutral ground in December 2025, when Anthropic donated MCP to the Agentic AI Foundation under the Linux Foundation, with OpenAI and Block as co-founders.

A second layer of the market has formed around the protocol's rough edges, in four rough categories: registries and marketplaces for discovery (the official MCP Registry, plus community catalogs); authentication providers implementing the spec's OAuth 2.1 evolution; gateways and proxies that centralize connections, credentials, and logging for enterprises; and hosting/runtime platforms that run servers so teams don't self-host. The spec has matured too — OAuth 2.1 as the auth foundation, resource indicators against token misuse, streamable HTTP for remote servers, async tasks, and sandboxed MCP Apps.

Two facts frame the rest of this paper. First, enterprises are deploying now: 41% report production MCP use, with another 30% piloting. Second, the protocol's stewards openly list what is still missing for the enterprise — standardized audit trails, multi-tenancy isolation, rate limiting and cost attribution, and defined gateway authorization behavior are roadmap items, expected largely as extensions. Production adoption is running ahead of production governance. That gap is the market.

2 · The documented reality: connection without protection

The security research community has now measured the MCP ecosystem at scale, and the results would be disqualifying in any other enterprise software category:

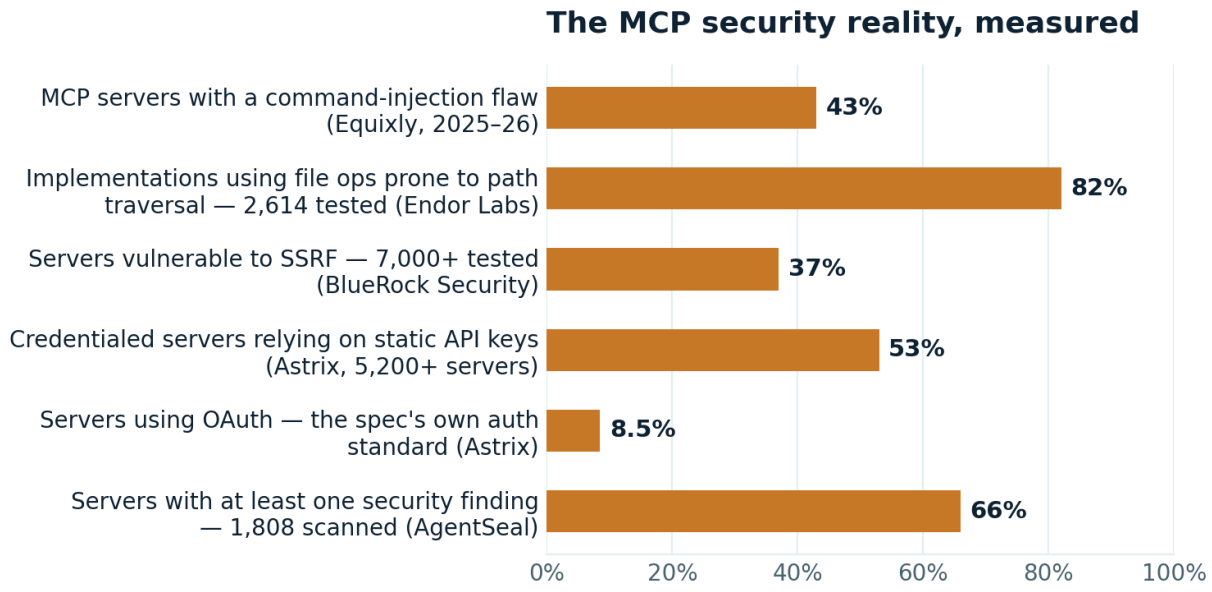

Figure 1 — Published scan and audit results across the public MCP ecosystem. Sources: Equixly; Endor Labs; BlueRock Security; Astrix; AgentSeal (via Practical DevSecOps' 2026 MCP security statistics compilation).

- Exposed infrastructure: Trend Micro found 492 MCP servers listening on the public internet with no authentication or encryption; a follow-up scan found the count had nearly tripled to 1,467.
- Live exploitation: recent CVEs include mcp-remote (CVSS 9.6, 437,000+ affected downloads), MCP Inspector RCE via DNS rebinding (CVSS 9.4), and 'MCPwn' (CVSS 9.8) — the last reported as actively exploited.
- Novel attack classes: tool poisoning (malicious instructions hidden in tool descriptions), rug pulls (a benign server updated into a malicious one), and confused-deputy flows are documented against real servers; an academic study found tool poisoning in ~5.5% of 1,899 scanned servers.
- The incidents have started: 88% of organizations in one 2026 survey reported confirmed or suspected AI-agent incidents; HiddenLayer attributes more than one in eight reported AI breaches to autonomous agents; Gartner projects a quarter of enterprise GenAI applications will suffer repeated security incidents by 2028.
- And governance is not keeping up: only 23% of organizations report a formal agent identity strategy, and one analysis finds just 14.4% of agents reach production with full security approval.

None of this means MCP was badly designed. It means MCP was scoped as a connection protocol while the market deployed it as an operations layer. The vulnerability statistics are the visible symptom; the structural gaps beneath them are the disease.

3 · POV: the five gaps between a plug and a platform

Gap 1 — Identity and authority: a connected tool is an authorized tool

MCP's trust model is binary: once a server is configured, its tools are available, typically riding a static credential with the union of every permission the integration might ever need. The Astrix audit quantifies it — 53% of credentialed servers on static keys, 79% passing secrets through environment variables, OAuth at 8.5%. There is no native notion of per-action authority: who may invoke this tool, against which resource, at what risk, for how long, revocable by whom. Every serious enterprise access-control principle of the last decade — least privilege, just-in-time credentials, scoped grants — is absent at exactly the layer where an autonomous agent acts.

Gap 2 — No verification: 'done' is whatever the server says

An MCP tool call returns a result; nothing in the ecosystem checks that the claimed effect actually occurred, touched only the intended target, and left no side effects. Agents chain dozens of calls on unverified claims — which is how a single poisoned tool description or hallucinated parameter propagates into real systems.

Gap 3 — No audit standard: the spec's own roadmap says so

When the auditor asks 'what did AI do in our systems last quarter, and who allowed it?', today's MCP estate answers with scattered client logs at best. Standardized audit trails, multi-tenancy isolation, and cost attribution are explicitly open items on the protocol's enterprise roadmap. Regulated industries cannot run consequential automation on infrastructure whose audit story is a to-do list.

Gap 4 — Supply-chain trust: a registry is not a vetting

Ten thousand public servers, most community-authored, installed by convenience: the scans above show what rides in. Tool poisoning, typosquatting, and rug-pull updates are supply-chain attacks with an agent — not a developer — as the target, and the current answer ('read the code yourself') does not scale to an enterprise estate.

Gap 5 — Gateways manage connections, not accountability

The emerging MCP gateway category is genuinely useful — centralizing credentials, allowlists, and logging. But even a well-run gateway answers 'which servers may we reach?' It does not answer 'was this specific action authorized, did it verifiably work, and can we replay it for the regulator?' Connection management is necessary. It is not governance.

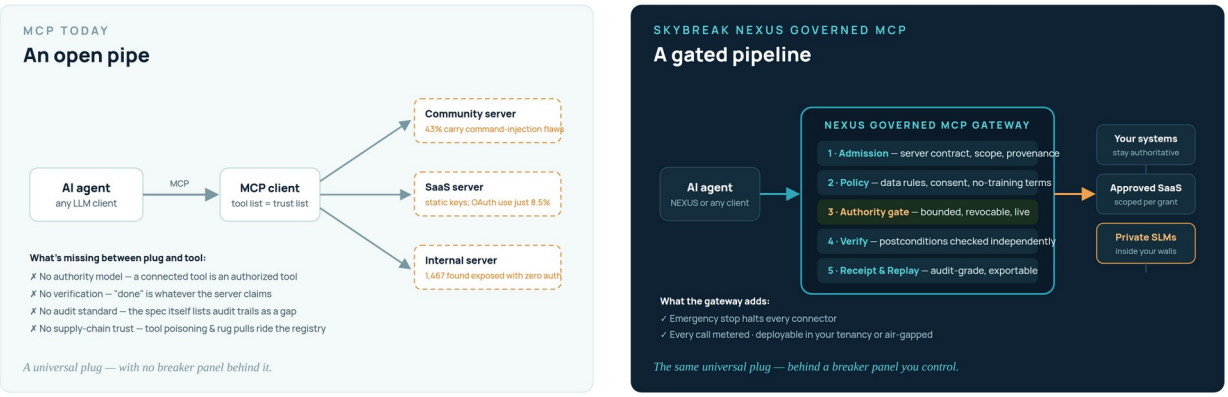

Figure 2 — The open pipe vs. the gated pipeline. Left-panel figures: Equixly (command injection); Astrix (static keys, OAuth); Trend Micro (exposed servers); protocol roadmap (audit gap).

4 · The SkyBreak answer: Governed MCP

NEXUS approaches MCP the way it approaches every capability: keep the open standard, add the governance the enterprise actually needs. The NEXUS Governed MCP gateway is MCP-compatible on both sides — existing clients and servers work unchanged — while every call now traverses five stages the open ecosystem lacks:

- Admission — no server joins by download. Each connector is admitted under a contract capturing identity, provenance, declared capabilities, scope, health, and receipt obligations; the registry problem becomes a vetting discipline.
- Policy — data rules are enforced at the gateway: what categories of data may flow to which admitted server, consent and residency constraints, and no-training terms flowed down to any external model provider, with every exchange logged.
- Authority — every consequential call requires a live, bounded, revocable Authority Grant: actor, capability, target, duration, risk tier. Static keys give way to scoped, expiring authority; autonomy is graduated and earned. A recommendation is never permission.
- Verification — postconditions are checked independently of the tool's own claim before an action counts as done; failures don't silently propagate down an agent chain.
- Receipt & Replay — every call yields an audit-grade record (who, what, when, under whose authority, with what verified result), retained and replayable. The audit trail the spec hopes to standardize someday is the system of record today — and it is the same Receipts fabric that meters billing, with no fees on failed or cancelled calls.

Around the pipeline sit the enterprise controls: a human-held emergency stop that halts every connector; deployment inside the client's own boundary — cloud tenancy, data center, sovereign region, or air-gapped — so no SkyBreak-operated server listens on the public internet; and enterprise-specific private SLMs doing the reasoning, so the intelligence consuming the tools is as private as the pipeline governing them.

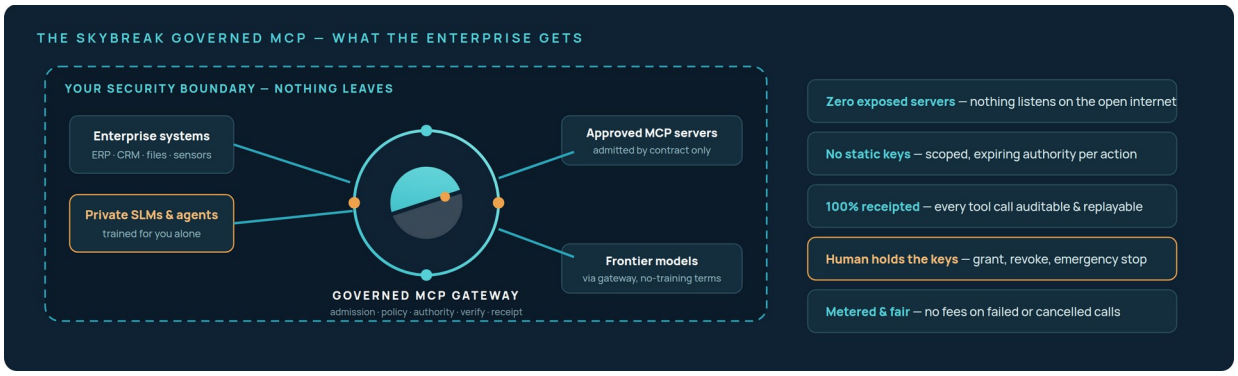

Figure 3 — The SkyBreak Governed MCP: everything inside the client's boundary, external reach only by admitted contract, value guarantees on the right.

4.1 Why this is next-level, not incremental

Map the five documented gaps to the five stages and the claim writes itself: identity and authority gaps are closed by Authority Grants; the verification gap by independent postcondition checks; the audit gap by Receipts and Replay; the supply-chain gap by contractual admission; and the gateway gap by the fact that

this gateway governs actions, not just connections. Security improves not by adding a scanner to a risky architecture but by removing the risky architecture: no static keys to steal, no exposed servers to find, no unvetted tools to poison, no unverified 'done' to trust. And because the governance layer is the same NEXUS spine that runs Missions, an enterprise that starts with Governed MCP is one step from governed operations end to end — the full closed loop documented in our companion papers.

> **Open MCP asks the enterprise to trust ten thousand strangers' servers. Gateways ask it to trust a connection list. SkyBreak asks it to trust nothing — and verify everything, with receipts.**

4.2 What the enterprise gets

- A safe on-ramp to the whole MCP ecosystem — the reach of the open standard without inheriting its measured 43%/82%/8.5% risk profile.
- Audit-ready AI operations from day one: every agent action reconstructable for security, compliance, and regulators.
- Credential hygiene by architecture: scoped, expiring authority replaces the static-key sprawl that dominates today's estate.
- Private by default: gateway and SLMs in your boundary; zero public listeners; data residency by configuration.
- Economics that match value: metered governed calls with fairness rules, creditable paid-proof entry, and SkyBreak's priced ≠ proven discipline — capabilities are sold at the maturity rung they have actually earned.

5 · Conclusion

MCP won. The plug is universal, the ecosystem is vast, and the enterprise is already committed — 41% in production while the protocol's own roadmap still lists audit, isolation, and authorization as open questions, and while measured studies find the server ecosystem riddled with basic flaws. That combination — maximum adoption, minimum governance — is precisely the gap SkyBreak was built for. The NEXUS Governed MCP keeps everything the market loves about the standard and adds the one thing the market cannot ship as a spec extension: accountability. Admission instead of installation, authority instead of keys, verification instead of trust, receipts instead of logs, and a human holding the stop button — privately deployed, on your data, with your SLMs.

Bring us your MCP estate — or the one your teams are building without telling you. We'll put a breaker panel behind it, as a fixed-price paid proof with gates, and hand you the receipts. info@skybreak.us

Sources

Market and security figures draw on the following public sources (accessed September 2026):

- Digital Applied, 'MCP Adoption Statistics 2026' — ~10,000 active public servers and 97M monthly SDK downloads (per Anthropic's December 2025 ecosystem announcement); 9,652 official registry entries; 15,926 GitHub mcp-server repositories; Stacklok survey: 41% production / 30% pilot; first-party vendor support list. digitalapplied.com.
- WorkOS, 'Everything your team needs to know about MCP in 2026' — protocol history and spec evolution (OAuth 2.1, resource indicators, CIMD, streamable HTTP, async tasks, MCP Apps); AAIF/Linux Foundation donation with OpenAI and Block; registry launch and growth; named client/server ecosystem; enterprise gaps acknowledged on the roadmap: standardized audit trails, multi-tenancy isolation, rate limiting/cost attribution, gateway authorization behavior. workos.com.
- Practical DevSecOps, 'MCP Security Statistics 2026' (compilation with named primary sources) — Equixly: 43% of tested servers with command injection, 30% SSRF; Endor Labs: 82% of 2,614 implementations with path-traversal-prone file operations; BlueRock Security: 36.7% SSRF across 7,000+ servers; Enkrypt AI: 33% of 1,000 servers with critical vulnerabilities; Hasan et al.: ~5.5% tool poisoning across 1,899 servers; AgentSeal: 66% of 1,808 servers with ≥1 finding; Astrix: 53% static keys, 8.5% OAuth, 79% env-var secrets across 5,200+ servers; Trend Micro: 492→1,467 exposed servers; CVE-2025-6514, CVE-2025-49596, CVE-2026-33032; Gravitee: 88% reporting agent incidents; HiddenLayer: >12% of AI breaches from autonomous agents; Gartner 2028 incident forecast; 23% agent-identity strategy; 14.4% full security approval. practical-devsecops.com.
- SkyBreak — the NEXUS Governed MCP description, five-stage pipeline (admission, policy, authority, verification, receipts/replay), private deployment model, SLM exclusivity, metering fairness, and truth-ladder discipline are SkyBreak's own positioning; forward-looking capability claims are subject to SkyBreak's paid-proof gates (priced ≠ proven). Companion papers: 'The Trust Deficit,' 'The Weakest Link,' 'The Empty Golden Record' (September 2026).