DATASHEET

MCP SECURITY · AI GOVERNANCE · VISIBILITY

YOUR AI AGENTS ARE CALLING PRODUCTION SYSTEMS RIGHT NOW. ON WHOSE AUTHORITY?

10,000+ **active public MCP servers, with adoption accelerating nearly 1,000x in eighteen months. Each one is a privileged path bridging an AI model into databases, code repositories, and internal APIs, provisioned faster than any governance framework can track, per Anthropic's MCP Ecosystem Update, December 2025.**

AI agents and MCP servers reach your production systems autonomously, every minute of every day. They hold credentials no one formally issued, invoke tools no one approved, and leave a trail no one can reconstruct. The market has noticed, every identity vendor now claims an answer. The question is no longer whether agent governance exists. It is whether it reaches every identity in your estate, from one place, before any call executes. Whiteswan governs at the chokepoint, across the whole estate, additive to the stack you already run.

OWASP Agentic Top 10 Aligned | AI Agents & NHI · MCP Server Governance | Per-Call Inspection

WHITESWAN
IDENTITY SECURITY

The Problem

THREE GAPS IN YOUR ENVIRONMENT NO MATTER HOW MANY VENDORS CLAIM THE CATEGORY

The market filled with AI-agent-security messaging faster than enterprises filled the actual gaps. In your environment, three remain open right now, and a louder market has not closed any of them.

Three open right now

1. MCP Security · the urgent gap

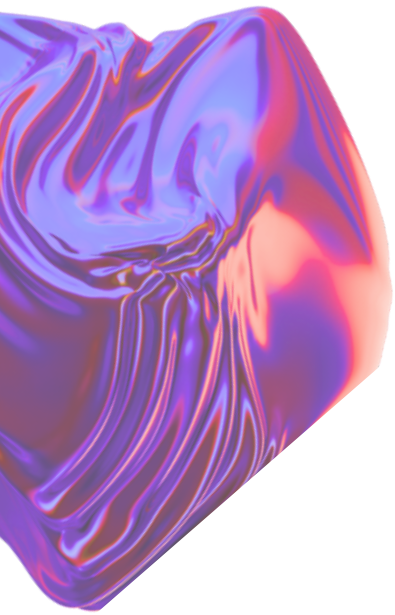
MCP servers bridge AI models directly into databases, CRMs, code repositories, and internal APIs. Every server is an ungoverned privileged access path. They are deployed by developers and AI teams faster than security can inventory them, and a governance tool that lives in a platform you have not deployed, or a cloud you do not run, does not see them.

2. AI Governance · the accountability gap

The EU AI Act, emerging AI disclosure rules, and internal AI policies all demand the same thing: a documented authorization chain and audit trail for every AI system touching production data. Owning a vendor's agent module is not the same as having that chain. When a regulator asks who approved an agent and what it was permitted to do, the answer has to exist as evidence, not as a roadmap line.

3. Visibility · the discovery gap

You cannot govern what you cannot see. AI agents and MCP servers are running today that were never formally provisioned, never inventoried, and never connected to any security tool. Cloud-native tools see the cloud. Auth-layer tools see the login. The agents reaching your legacy and on-prem systems, the estate the newest tools skip, stay invisible.



April 25, 2026

PocketOS

An AI agent deleted a production database and all backups in 9 seconds. It found an API token in an unrelated file, decided to delete a volume to fix a credential mismatch, and issued one API call. No confirmation. No scope check. No audit trail. None of the five questions above could be answered.

"Publicly disclosed by PocketOS founder Jer Crane"

Platform Capabilities

EVERY AGENT CALL.
GOVERNED BEFORE IT EXECUTES.

Whiteswan governs every AI agent and MCP server at the protocol chokepoint, the point every call must pass through to reach an enterprise system. Discovery, policy, inspection, and revocation happen there, through a single policy engine, before any action lands.

Surface	Security Challenge	Risk	Whiteswan Control
MCP Servers	Ungoverned privileged path into core systems	Critical	MCP lifecycle governance at the chokepoint
AI Agents	Standing credentials, no approval, no scope	Critical	Per-call inspection · JIT agent policies
Agent → Tool calls	Unregistered tools invoked without control	High	Capability allowlisting · pre-execution check
Delegated access	No record of who authorized the action	High	Token classification · human-initiator trace

MCP SECURITY · AI GOVERNANCE · VISIBILITY

MCP Lifecycle Governance

Full lifecycle for every MCP server reaching your systems: Discover, Govern, JIT, Monitor, Revoke: enforced at the protocol chokepoint. Allowlist controls which servers agents may connect to; unregistered servers are uncallable.

Per-Call Inspection: 10-Step Orchestrator

Every agent call passes a 10-step inspection sequence before execution: identity validated against your IdP, capability checked against the allowlist, scope verified, call logged. No exceptions, no bypass.

Capability Allowlisting & Agent Approval

Unregistered capabilities are uncallable. Agents are approved at the deployment level, with an admin moving each agent from PENDING to APPROVED or BLOCKED. Per-tool scope policy is expressible via Cedar PDP; per-tool UI approval is in development.

Cryptographic Agent Identity

Per-session cryptographic identity issued via SPIFFE/SPIRE at agent spawn. Every call is validated against your IdP via JWT or SHA-256 hashed API key (wsg_ prefix). Orphan credentials never issued through the governed workflow are blocked at the gateway.

Token Classification & Delegation Lineage

Token classification distinguishes delegated tokens (on-behalf-of) from automated agent credentials, naming the original human initiator across the full chain. Multi-hop delegation is Verified via RFC 8693 Token Exchange at every hop, human to tool.

JIT Access: Per Agent, Per Call

No standing admin keys for any AI agent, ever. Credentials are issued for the duration of a specific task and revoked automatically on completion. The lateral movement paths standing privileges create are structurally eliminated, not monitored.

Governance lifecycle (horizontal flow):

Discover — Find every agent & MCP at the chokepoint → **Govern** — Apply policy & allowlist → **JIT Access** — Time-bound per-call → **Monitor** — In-flight inspection → **Revoke** — Auto - On Task Completion

Chokepoint discovery surfaces every agent and MCP server that routes through the Whiteswan gateway. Discovery of shadow agents in un-instrumented environments is extended via a network sensor or CNAPP layer.

Platform Capabilities

Beyond the Tool Call: Agent-to-Agent Traffic, Data Leaving the Gateway, and Provable Posture

MCP governs the tool call. As agents increasingly hand work to other agents, and as every tool response carries the risk of exposing something sensitive, visibility has to extend past that single call. These four capabilities close that gap and turn the audit trail into evidence a CISO or auditor can act on.

Governance Surface	Security Challenge	Risk	Whiteswan Solution
Agent-to-Agent Handoffs	No policy check between agents, only at the tool call	Critical	A2A Protocol Governance
Sensitive Data in Responses	Tool outputs returned with no inspection	Critical	Egress Data Classification
Credentials at Rest	MCP credentials stored in plaintext	High	AES-GCM Encryption
Audit-to-Compliance Mapping	Logs exist, but not mapped to a framework	High	CISO Governance Posture · Control Register

MCP Gateway
Governance Beyond the Chokepoint

Agent-to-Agent (A2A) Governance

The gateway evaluates A2A protocol traffic with the same approve-before-connect policy that governs MCP tool calls. Whether an agent calls a tool or another agent, the decision happens at the same chokepoint, same audit trail.

Egress Data Classification & Cred Encryption

Every response is scanned for sensitive data (Luhn-checked cards, IBAN mod-97, structural SSN, JWTs, PEM keys, labelled secrets) before it leaves. MCP credentials, headers, URL parameters, and config are encrypted with AES-GCM at rest.

CISO Governance Posture

Gateway telemetry rolls into a graded, explainable posture score: governed requests, sensitive events, policy gaps, blocked/denied actions, active agents, and risky principals, tracked over 24h/7d/30d/90d, with priority actions assigned to named owners.

Control Register

Gateway audit data maps directly to SOC 2 Trust Services Criteria: CC6.1, CC6.2, CC6.3, CC7.2, CC7.3, and CC8.1, each individually labeled Evidenced. Not a certification claim, the enforcement itself becomes the evidence a regulator asks for.

Proof lifecycle (horizontal flow):

Detect — A2A call or response in transit → **Classify** — Score against policy & sensitivity → **Enforce** — Allow, deny, or redact → **Score** — Roll into posture grade → **Evidence** — Map to Control Register

Whiteswan aligns to SOX, PCI-DSS, ISO 27001, SOC 2, HIPAA & GDPR frameworks.

Governance & Accountability

A GOVERNANCE CHAIN A REGULATOR CAN READ

Whiteswan turns every agent action into a governed, attributable, auditable event: the authorization chain a regulator asks for, existing as evidence rather than intent. A live registry of every agent and MCP server, an immutable trail across 48 event types, and the original human initiator named on every call.

Code	Risk	Coverage	Whiteswan Mechanism
ASI01	Agent Goal Hijack	Partial	JIT access limits blast radius: a hijacked agent cannot act beyond its task-scoped, time-bound credentials
ASI02	Tool Misuse and Exploitation	Full	Cedar policy enforces PERMIT/DENY before execution. Token classification identifies the original human initiator across the delegation chain.
ASI03	Identity and Privilege Abuse	Full	Cryptographic session identity. Confused deputy attack closed. JIT revocation eliminates cached credential escalation.
ASI04	Agentic Supply Chain	Partial	MCP Gateway enforces allowlist; unauthorized server connections are blocked. Upstream integrity is not in scope.
ASI05	Unexpected Code Execution	Out of scope	LLM-layer concern. Whiteswan governs agent credentials, not code generation.
ASI06	Memory and Context Poisoning	Out of scope	LLM-layer concern. Requires dedicated LLM security tooling.
ASI07	Insecure Inter-Agent Communication	Partial	Only agents with valid session credentials join a governed workflow. Transport security not in scope.
ASI08	Cascading Failures	Partial	JIT limits cascade depth structurally. A compromised agent cannot trigger unrestricted downstream actions beyond its task-scoped credentials.
ASI09	Human-Agent Trust Exploitation	Out of scope	HITL gates create checkpoints. Cannot prevent human deception by AI-fabricated justification.
ASI10	Rogue Agents	Partial	Identity inventory detects unauthorized spawning. Agentic Gateway credential-starves ungoverned agents.

All governance events are logged across 48 event types and exportable via API. Pre-built, regulator-shaped report templates are a reporting-layer build: the data is complete, the formatted export is not yet.

Coverage key:

Full: direct, defensible coverage
Partial: identity and access layer addressed; full risk requires complementary tooling
Out of scope: LLM reasoning layer; requires dedicated AI security tooling

Whiteswan aligns to SOX, PCI-DSS, ISO 27001, SOC 2, HIPAA & GDPR frameworks.

Where Whiteswan Sits

GOVERNING AGENTS IS NOW TABLE STAKES

GOVERNING THE WHOLE ESTATE, FROM ONE ENGINE, ISN'T.

Capability	Platform Extension	Auth-Layer Inline	CSP Native Tools	Whiteswan
Governs AI agents & MCP servers	✓	✓	✓	✓
Post-authentication, per-call inspection at the chokepoint	Partial	Auth-layer only	Partial	✓
Legacy on-prem + AD (Windows 2003–2022) in the same engine	Partial	✗	✗	✓
Cloud NHI + agentic from one engine, no separate module	Separate module	✗	Cloud-only	✓
Additive, no sensor, no rip-and-replace	Varies	✓	Varies	✓
No platform or suite prerequisite to adopt	✗	✓	✓	✓
Time to first value:	Weeks to months	Weeks	Weeks	Days, not weeks
Honest live-vs-roadmap claim perimeter	Varies	Varies	Varies	✓

AI agent governance is no longer an empty field. Suites, inline tools, and cloud-native point tools all reach for it. What none of them combine: post-authentication control at the chokepoint, across legacy on-prem, cloud, and agentic from one engine, additive, in days not weeks, with an honest line between what ships and what's in design.

Deployment details

⚡ Pilot

A complete inventory of every AI agent and MCP server reaching your systems, capability allowlisting and per-call inspection active across the highest-risk workflows, per-call validation against your IdP, and the first immutable audit trail established, in days not weeks. No infrastructure overhaul, no platform prerequisite.

🔧 Hybrid Enforcement

Lightweight agents on Windows and Linux. Agentless gateways available for constrained environments. Sub-100ms access decisions.

⚙ One Engine, Whole Estate

On-prem, cloud, and AI agent governance from a single engine, covering legacy AD through agentic, AWS · Azure · GCP. One place to write policy, one audit trail. Not a module on a suite you must own first.

🔌 Native Integrations

Works alongside the SIEM, XDR, and endpoint tooling you already run, enhancing your existing investment. Whiteswan is additive to every deal, enhancing, not replacing, your existing security stack.

Whiteswan is not the only way to govern an agent. It is the clearest, and the one that governs your whole estate from one place.

WHITESWAN
IDENTITY SECURITY

Recognised for AI-driven identity security and multi-cloud integration. KuppingerCole Rising Star, October 2025.

"Whiteswan gave us visibility into identities we didn't know were exposed." - Moosa M, Exotel

The deployment was faster than any PAM project we had attempted before." - Kamalesh Kumar, MG Contractors

"The lateral movement containment capability was exactly what our board was asking for." - Puneet Sharma, Rockman Industries

EVERY AI AGENT. EVERY MCP SERVER. ONE POLICY ENGINE.

AI agents and MCP servers carry admin-level access to production systems in every enterprise. Whiteswan governs that surface, from credential issuance through real-time inspection through automatic revocation, from a single policy engine.

- **NHI Risk Assessment:** Discover every ungoverned AI agent and MCP server, including the ones your existing tools cannot see.
- **Agent Governance Pilot:** capability allowlisting and per-call inspection live, per-call IdP validation active, first immutable audit trail established in days/weeks.
- **ROI Analysis:** Calculate the risk reduction value of governing every AI agent credential in your environment.

vmamidi@whiteswansecurity.com

www.whiteswansecurity.com