Whitepaper

MORPHISEC

Governing the Autonomous Enterprise

Closing the AI Security Gap at the Point of Execution

Why shadow AI and compromised agents demand endpoint-native prevention – not another detection layer.

Authored by Brad LaPorte,
Morphisec CMO and Gartner Veteran

Table of Contents

01 | Executive Brief: The New AI Risk Lives on the Endpoint

Artificial intelligence is no longer emerging technology.

It is embedded across the enterprise. From coding assistants and copilots to autonomous agents and API-driven workflows, organizations are integrating AI into daily operations at remarkable speed. By the end of 2026, an estimated 85% of enterprises will have deployed or planned AI agents in production, and 40% of enterprise applications will integrate AI agents, up from just 5% in 2025.

That adoption has opened a new, largely invisible attack surface, and it converges on a single place: the endpoint.

Employees run AI desktop apps and local LLMs outside IT oversight. Developers wire AI assistants and Model Context Protocol (MCP) connectors directly into their workflows. Autonomous agents execute multi-step tasks across systems with broad, inherited permissions. And attackers now wield the very same AI to move at machine speed.

Most enterprise security architecture was built for a different era; one defined by human-driven attacks and observable patterns, secured by detection-first logic: identify, alert, respond. That model is breaking down. AI-driven actions complete their objectives before detection tools can analyze them, and the AI running inside the enterprise sits in a blind spot that firewalls, identity, and EDR were never designed to govern.

Firewalls, identity, and EDR were built to watch people and devices.
None of them govern what AI can reach, execute, or weaponize on the endpoint.

The result is a structural problem we call the AI Security Gap – the growing inability to govern AI-driven activity, whether it originates from an external attacker or an employee's own tools. This paper defines that gap, explains why detection and network-based controls cannot close it, and lays out a prevention-first model for governing AI exactly where it acts: on the endpoint.

That model is AI Usage Control (AIUC) from Morphisec.

02 | The AI Security Gap: Visibility, Control, and Prevention

AI systems are not static applications.

They are dynamic, adaptive, and increasingly autonomous. They read data, make decisions, and take action across systems. That makes them fundamentally harder to secure than traditional software. The AI Security Gap is not a single vulnerability; it is a structural limitation defined by three core failures.

1. The Visibility Gap

Organizations cannot secure what they cannot see. In most environments today there is no reliable way to identify every AI tool and agent operating on endpoints, distinguish authorized from unauthorized use, or monitor how AI behaves in real time. Meanwhile those systems quietly access sensitive data, interact with source-code repositories, call external APIs, and execute automated workflows. The result is a fragmented, incomplete picture of the AI attack surface; and security teams operating blind.

2. The Control Gap

Even when AI usage is identified, organizations often lack the ability to control it. Perimeter defenses, network monitoring, and data-loss prevention were never designed to govern how AI behaves at runtime. Without enforcement at the endpoint, there is no way to define acceptable AI behavior, prevent unauthorized actions, or terminate a risky process in real time. AI systems are not just unmonitored. They are effectively unmanaged, raising the risk of data exfiltration through AI APIs, unauthorized system modification, privilege escalation, and supply-chain exposure.

3. The Prevention Gap

The most critical failure lies at the point of execution. Modern tools are designed to detect and respond, not to prevent before execution. In the context of AI, that is fatal: by the time suspicious behavior is identified, the action has already occurred. This is especially true at the endpoint, where AI processes run locally, memory-resident and fileless techniques evade inspection, encrypted channels obscure content, and actions complete faster than analysis cycles. Detection, by design, is reactive, and reactive security cannot keep pace with machine-speed execution.

Closing the gap requires a structural shift toward:

- **Continuous visibility** – Into AI tools, agents, and behaviors at the endpoint.
- **Runtime control** – Over how AI is allowed to operate within the environment.
- **Pre-execution prevention** – To stop malicious or unauthorized AI actions before they occur.

03 | Two Faces of AI Risk: Shadow AI and Compromised AI

The AI Security Gap manifests in two distinct but related ways. Both bypass legacy controls, and both act on the endpoint.

Shadow AI – Unmanaged AI, Outside IT View

Staff paste confidential data into AI assistants and run AI desktop apps and local LLMs outside IT oversight, touching files and workflows directly on the device. Development teams adopt AI coding tools and connect MCP servers without governance. None of it appears in a central inventory, and none of it is bound by policy.

Compromised AI – Trusted AI, then Hijacked

Enterprise AI agents run with broad, standing permissions. A poisoned prompt, a manipulated tool call, or a supply-chain compromise inherits that trust, and the malicious commands look completely legitimate. Because the activity occurs inside an approved application and workflow, it rarely trips a traditional alert.

The more AI an enterprise adopts, the larger its attack surface grows; the more attackers leverage AI, the more scalable their attacks become. Without endpoint-level governance, that loop accelerates risk exponentially.

51%
of enterprises run AI agents in production today

40%
of enterprise apps will integrate AI agents by end of 2026

44%
of all data breaches now involve ransomware

Sources: IDC / industry survey 2026; Gartner, Aug 2025; Verizon DBIR 2025.

04 | Why Network, CASB, and Browser Controls Miss AII

Faced with AI risk, many organizations reach for the tools they already own: network gateways, cloud access security brokers (CASB), and enterprise-browser or browser-extension controls. These help with sanctioned, routed SaaS traffic, but they share a structural blind spot: they only see what passes through them. AI increasingly does not.

- **Local LLMs and CLI agents** – Tools like Ollama and command-line AI assistants execute entirely on the device, never crossing a network gateway.
- **IDE-embedded and desktop AI** – Copilot, Cursor, and desktop assistants act inside trusted applications, invisible to a browser extension.
- **MCP connectors** – Agent-to-tool connections grant powerful, file-level and system-level reach that traffic inspection never sees.
- **Encrypted and air-gapped activity** – Encrypted API calls obscure content, and offline or air-gapped endpoints are entirely beyond network reach.

The field of AI-security tools effectively splits two ways, and each leaves the endpoint exposed. Network, SASE, CASB, and browser tools see only routed traffic or a single browser tab. Detection-and-respond endpoint platforms mostly classify and react after the fact, and require you to adopt their platform. Neither governs AI at the moment of execution on the agent you already run.

Capability	Network / CASB / Browser AI Tools	Endpoint-Native AI Governance Tools
Visibility	Routed traffic or browser tab only	Local LLMs, CLI agents, IDE & desktop AI, MCP
Enforcement point	After traffic leaves, or in the browser	At execution, on the endpoint
Action on risk	Detect, classify, log	Default-deny prevention – stops the action
New agent required	Often a proxy, gateway, or browser	None – rides the agent already deployed
Offline / air-gapped	Blind	Fully covered
Anti-ransomware tie-in	None	Same engine that already stops ransomware

05 | AI Threats Already in the Wild

These are not hypotheticals.

Across 2025, researchers documented AI being weaponized at every stage of the attack chain, and each case maps directly to the visibility, control, and prevention gaps above.

The Threat	How It Works	How Endpoint-Native Governance Stops It
PromptLock – polymorphic AI ransomware (ESET, Aug 2025)	Connects to a local Ollama endpoint and generates polymorphic Lua at runtime – unique code each run, evading static EDR.	Ollama auto-flagged as an AI agent; file-write spike + child-process spawn → Critical alert; local LLM blocked by policy.
QUIETVAULT – AI-weaponized supply chain (Google TI, 2025)	A compromised npm plugin weaponizes pre-installed AI CLIs to recursively scan and exfiltrate GitHub tokens and wallets.	SSH-key and credential-directory access blocked by Day-1 guardrails – no ML baseline required.
GTG-2002 – autonomous extortion (17 orgs)	An attacker runs Claude Code on a Kali host; the agent autonomously selects targets and sets six-figure ransoms.	Claude Code identified instantly; data-volume + sensitive-dir anomaly → High alert; egress policy blocks unapproved LLMs.
Unconstrained Deletion – runaway agent	A coding agent told to clean old records deletes the entire production database and all backups in under ten seconds.	Detects the MCP connector's DELETE access, blocks AI from backup locations, flags mass deletion – before it completes.

The common thread: the damage happened (or would have happened) on the endpoint, at machine speed, inside trusted tools. Only a control that sees and stops AI at execution could intervene in time.

06 | From Detection to Governance: Discover, Govern, Apply Guardrails, React

Closing the AI Security Gap requires moving security closer to where risk materializes (the point of execution), and pairing visibility with enforcement. Endpoint-native AI governance does this through four continuous moves.

- **Discover** – Continuously identify and inventory every AI tool, account, agent, browser extension, LLM service, and MCP connector – including shadow AI and unapproved tools.
- **Govern** – Map every AI action to user identity, AI service account, and device context, and decide which AI may operate – enforced by role or department.
- **Apply Guardrails** – Apply least-privilege policies that block risky AI actions at runtime: access to credentials and PII, writes to backup and recovery locations, and risky process spawns such as PsExec, vssadmin, and PowerShell.
- **React** – Give each AI tool its own local behavioral baseline and flag abnormal activity in real time – file-operation spikes, exfiltration behavior, abnormal connector usage, and suspicious execution chains.

> The question is no longer “How quickly can we detect this?” It is “Can we prevent this AI action from executing at all?”

Crucially, this is behavior control, not content inspection. Rather than reading prompts or decrypting data streams – which raises privacy and compliance concerns and still misses local execution – governance focuses on what the AI is doing. That makes it resilient to encryption, variability, and the legitimate-looking nature of AI activity.

07 | AI Usage Control: Endpoint-Native AI Governance

Morphisec AI Usage Control (AIUC) operationalizes this model. It is endpoint-native AI governance embedded directly in the Morphisec Protector across Windows, Linux, and macOS – with no new agent, no proxy, and no cloud relay.

Because it rides the agent you already deploy, AIUC sees and governs the CLI agents, local LLMs, IDE plugins, and desktop AI that network- and CASB-based tools structurally miss, including on offline and air-gapped endpoints. AIUC extends Morphisec's Anti-Ransomware Assurance Suite, powered by Automated Moving Target Defense (AMTD).

What makes AIUC different

- **Endpoint-native** – no new agent – Embedded in the Protector you already run. It sees AI where it actually executes: locally.
- **Full AI inventory** – Auto-discovers every AI tool, account, agent, extension, LLM service, and MCP connector – including shadow AI.
- **Prevention-first, tied to anti-ransomware** – Default-deny guardrails block risky AI actions with the same engine that already stops ransomware – before exfiltration or encryption. It stops the action, not just logs it.

What security teams gain

Outcome	What It Delivers
Full AI Inventory	Know every tool, user, and MCP connector – including everything IT never approved.
Ransomware Resilience	Block pre-encryption AI recon, backup deletion, LoLBin abuse, and credential theft before damage.
Compliance Evidence	Audit events, enforcement records, and inventory exports ready for EU AI Act, NIST AI RMF, ISO 42001, and SOC 2.
Safe AI Enablement	Let engineering use Copilot, Claude, and Cursor – guardrails target risky actions, not the tools.

AIUC fits within Morphisec's five-layer Anti-Ransomware Assurance Suite (Adaptive AI Defense, Adaptive Exposure Management, Infiltration Protection, Impact Protection, and Adaptive Recovery), operating on a continuous Predict → Prevent → Adapt loop, and backed by Morphisec's Ransomware-Free Guarantee.

08 | What Security Leaders Must Do Now

AI adoption is outpacing governance, and attackers are already exploiting the gap.

The question is not whether this shift will affect your organization, but how quickly you can adapt. Leaders should move deliberately on five fronts.

An endpoint-native AI governance checklist

- **Inventory AI at the endpoint** – discover every AI tool, agent, extension, and MCP connector – starting with shadow AI.
- **Tie AI actions to identity** – map each action to a user, service account, and device, and set policy by role.
- **Enforce least privilege at runtime** – block AI access to credentials, PII, and backups, and stop risky process spawns.
- **Baseline and watch behavior** – give each AI tool a local baseline and flag exfiltration and anomalies in real time.
- **Prove it for compliance** – generate audit-ready evidence for EU AI Act, NIST AI RMF, ISO 42001, and SOC 2.

Above all, enable rather than ban.

Default-deny guardrails should be policy-tunable and aimed at risky actions, not at the tools developers rely on. Done well, AI governance is safe enablement: engineering keeps Copilot, Claude, and Cursor, while the organization prevents harm.

09 | Conclusion: Govern the AI You Can't See

AI is redefining the enterprise, accelerating innovation while reshaping the threat landscape into something faster, more adaptive, and increasingly autonomous. The assumptions that guided a decade of detection-first security no longer hold: visibility is fragmented, control is limited, and the window for response is vanishing.

That is the AI Security Gap.

Closing it does not require another detection layer. It requires governing AI where it actually runs (on the endpoint) with continuous discovery, identity-aware control, least-privilege guardrails, and real-time behavioral response.

Morphisec AI Usage Control delivers exactly that, prevention-first and tied directly to anti-ransomware, on the agent you already run.

When AI accelerates the threat, preemption is peace of mind. Govern the AI you can't see – before it acts.

About Morphisec

Morphisec is the global leader in prevention-first cyber defense and anti-ransomware protection, delivering preemptive security that stops attacks before execution. Founded in 2014 and headquartered in New York, and powered by Automated Moving Target Defense (AMTD), Morphisec protects millions of endpoints across finance, healthcare, manufacturing, and technology. At Morphisec, we don't just respond – we prevent.

To learn more, visit morphisec.com/demo