WHITEPAPER

Netzilo

secure digital workplace for

# The Post-AI Enterprise

The modern digital enterprise is the result of a long and continuous evolution—and so are the security architectures built to protect it. In the early days of the internet, enterprise security was defined by a clear and defensible digital perimeter. Corporate applications and data resided behind firewalls, access was limited to trusted networks, and security was largely a matter of keeping threats out. As long as data stayed inside the perimeter, control was straightforward.

This model began to erode as enterprises were forced to extend access beyond the traditional boundary. Remote work, mobile users, and the rapid proliferation of SaaS applications pushed data outside the firewall. Virtual private networks (VPNs) and identity-based controls e.g. zero trust network access control (ZTNA) emerged to bridge this gap, and Zero Trust architectures became the new gold standard. As enterprise infrastructure expanded into private clouds and distributed data centers, security strategies increasingly shifted toward data-centric protection—focusing on securing data wherever it lived or traveled, rather than defending a fixed network edge. Today, that evolution has reached a new inflection point. The enterprise edge no longer stops at data centers, managed devices, or trusted SaaS providers. It now extends all the way to the endpoint—and beyond it. Wherever enterprise data is accessed, processed, consumed, or regenerated that location effectively becomes part of the enterprise boundary. This includes unmanaged devices, personal environments, and third-party execution contexts such as AI systems where traditional controls offer little to no visibility.

The rapid proliferation of generative AI has fundamentally amplified this challenge. The accelerated adoption of AI-based tools—ranging from external GenAI providers to AI-powered IDEs such as Cursor, and autonomous desktop AI agents—has created an entirely new class of risk. Enterprise data is now routinely delivered to systems that were never designed to be trusted, governed, or controlled. These tools introduce novel data exposure paths, often operating outside established security, compliance, and monitoring frameworks.

Today, a growing share of internet traffic is generated by AI systems rather than humans. AI agents have effectively become new operators of enterprise data—reading, transforming, generating, and redistributing information at scale. Generative AI is not simply another productivity tool. It is a transformative force that is reshaping how data is created, accessed, and consumed—and even how organizations are structured and managed. Decisions, workflows, and data interactions increasingly occur at the “speed of thought,” leaving most enterprises struggling to adapt their security posture quickly enough. As a result, many organizations are falling behind in securing what can now be described as the post-AI enterprise.

The post-AI enterprise demands a rethinking of the digital workplace security model—one that assumes the enterprise boundary is fluid, data is constantly in motion, and AI agents are integral to daily operations. Securing this new reality requires security architectures that are built explicitly for AI-driven data consumption, not retrofitted from pre-AI assumptions.

# Netzilo AI Edge

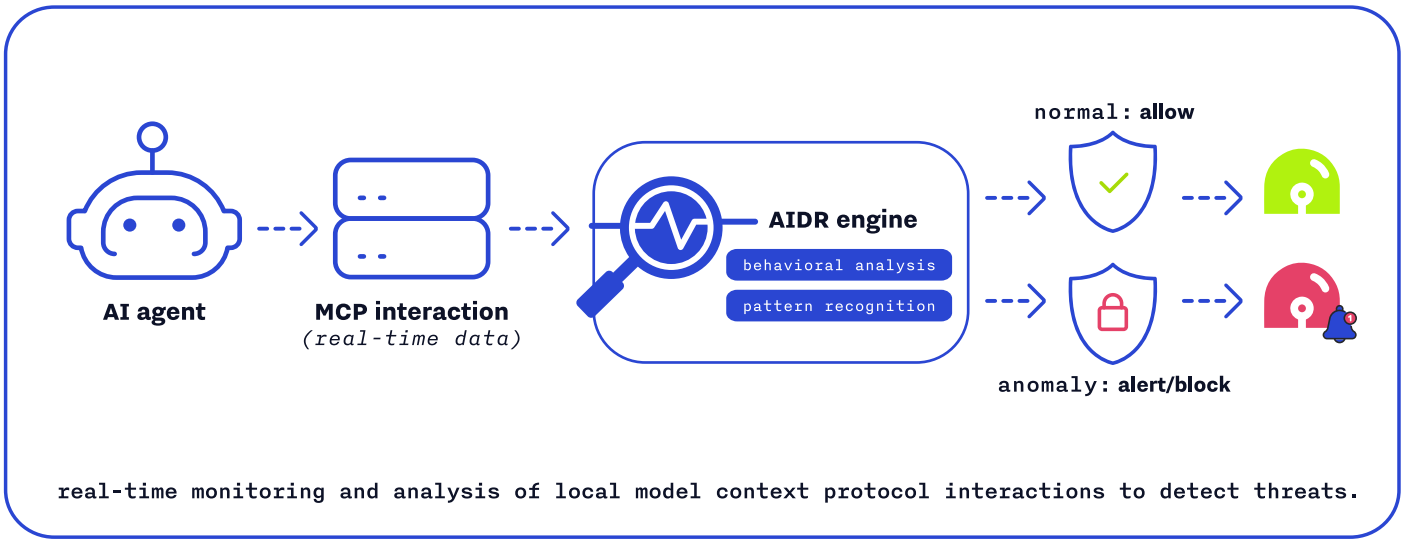
The browser, as we have known it for decades, is approaching obsolescence. In its place, AI agents are emerging as the primary interface between users, applications, and data. At the same time, the traditional SaaS consumption model is undergoing a fundamental transformation. Web applications are increasingly being re-architected as Model Context Protocol (MCP) servers; built not for direct human interaction, but for programmatic, agent- driven access.

This shift introduces an entirely new attack surface. Agentic workflows are vulnerable to novel threats such as prompt injection attacks, malicious or compromised MCP servers, tool poisoning, and data exfiltration through covert or non-obvious channels. These threats do not map cleanly to existing browser, network, or cloud security controls. Netzilo AI Edge is purpose-built for this new threat landscape, delivering a multi-layered defense architecture designed to secure AI-native workflows from the ground up.

## AI detection and response (AIDR)

Similar in concept to Endpoint Detection and Response (EDR), Netzilo AI Edge introduces AI Detection and Response (AIDR)—an endpoint-resident security layer specifically designed to monitor and protect AI agent activity. Unlike remote gateways or cloud- based inspection solutions that suffer from latency, blind spots, and incomplete context, AIDR operates directly on the endpoint where AI agents execute and where attacks originate.

AIDR continuously monitors local MCP interactions in real time, including stdio-based MCP servers such as filesystem tools and Node.js-based agent utilities. By observing behavior at the point of execution, Netzilo AI Edge can detect and respond to attacks that would otherwise remain invisible to traditional controls.



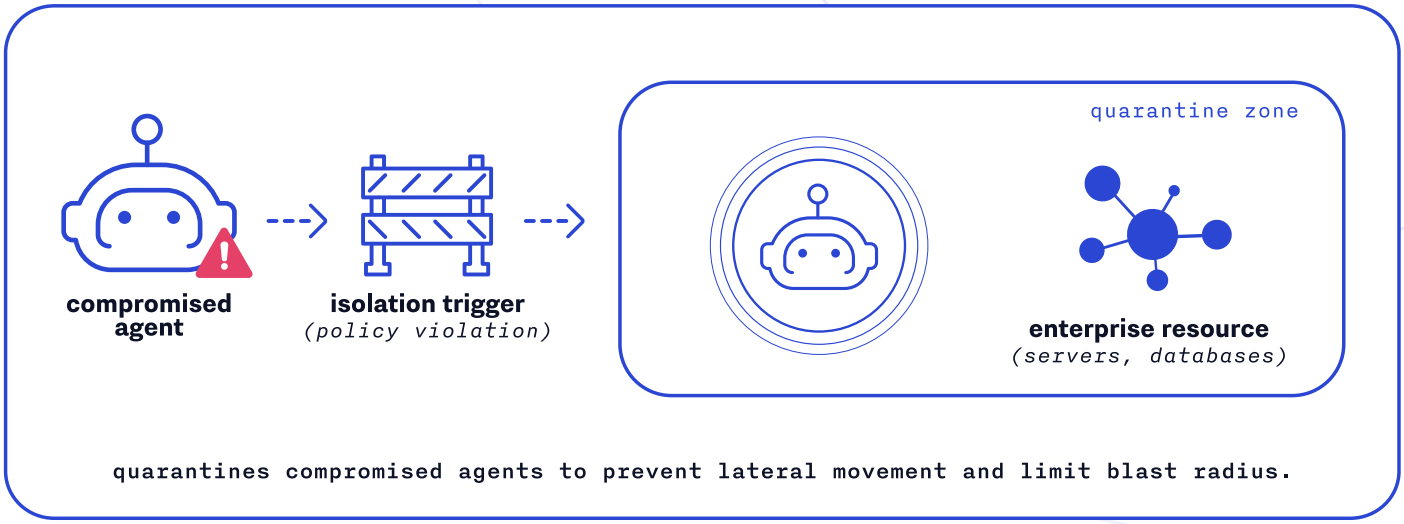

real-time monitoring and analysis of local model context protocol interactions to detect threats.

### key capabilities include:

- Real-time behavioral analysis using both static and dynamic detection techniques
- Interception and analysis of tool-call chains, including advanced policy enforcement such as Meta's Rule of Two
- MCP server reputation analysis, with the ability to restrict usage to approved or explicitly whitelisted MCP servers

This layer provides immediate visibility and control over AI agent behavior before malicious actions can propagate.

## AI agent isolation




Browser isolation has long been used as a secondary defense layer to contain threats originating from untrusted web content. However, this model has not kept pace with the emergence of AI agents, which now function as "the new browsers" of the enterprise.

Netzilo Workspace extends isolation principles to AI agents themselves. Using local isolation technology, AI agents are executed in a controlled, quarantined environment—segmented from the host system and the broader enterprise network. This approach prevents lateral movement, blocks privilege escalation, and dramatically reduces the blast radius of any successful compromise.

By isolating agent execution without sacrificing performance or usability, Netzilo AI Edge introduces a critical containment layer for agentic workloads.

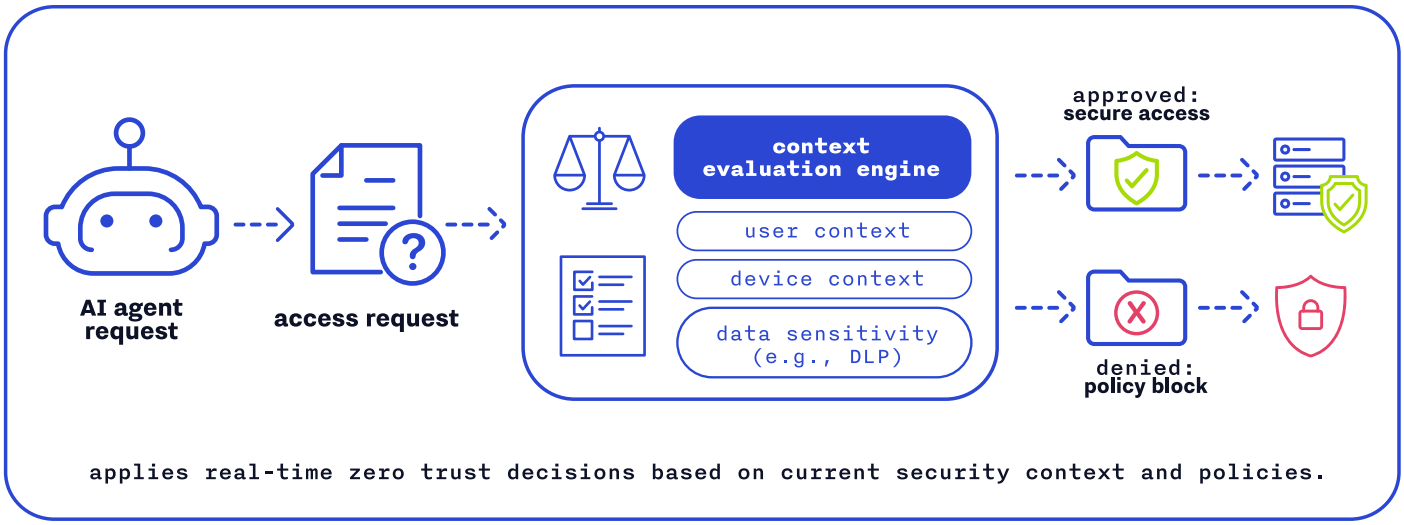
## endpoint security posture enforcement

Netzilo AI Edge enforces endpoint security posture on every AI agent action, enabling fine- grained, zero-trust decision-making at runtime. Security policies are dynamically evaluated based on the agent's context, tool usage, and overall trust posture.

This enables enforcement of questions such as:

- Can an AI agent download files from corporate SharePoint if Data Loss Prevention (DLP) controls are disabled?
- Can an AI agent access corporate MCP servers while simultaneously using unsanctioned or untrusted MCP tools?
- Can an AI agent interact with GitHub code repositories when non-approved MCP servers are active in its execution chain?

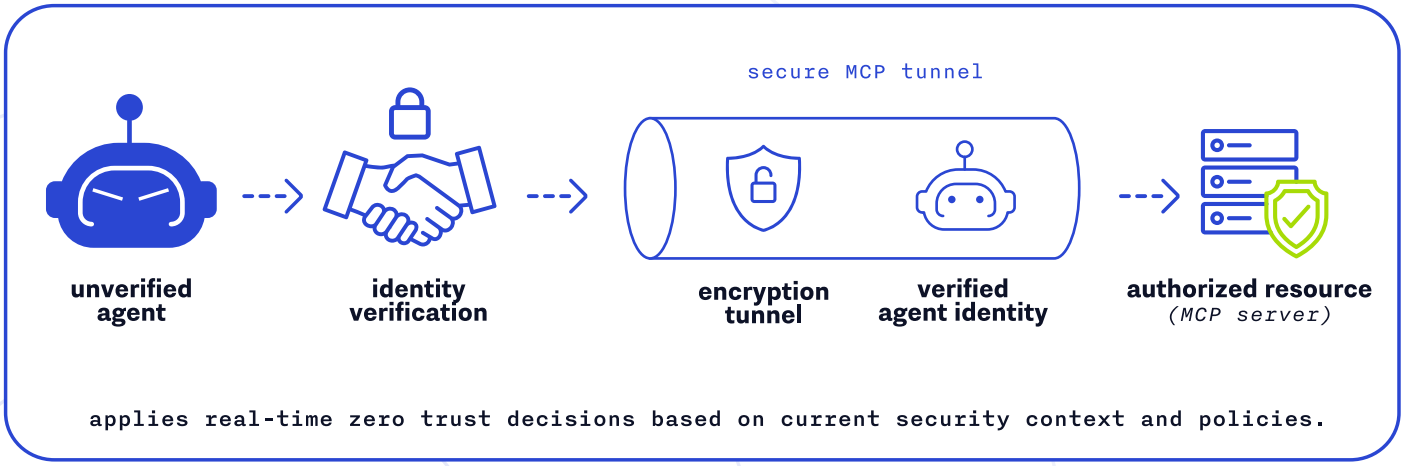
These controls allow enterprises to move beyond static allowlists and coarse-grained policies, enforcing context-aware zero trust for AI-driven operations.




## zero trust access & networking

In the post-AI enterprise, secure and reliable networking can no longer be limited to human users and applications. AI agents and MCP servers require the same—or stronger—access controls, identity guarantees, and policy enforcement traditionally reserved for users. Netzilo AI Edge extends Zero Trust networking principles to agent-to-agent and agent-to-MCP communications, delivering a fully integrated access layer for AI-native environments.

Just as modern Zero Trust platforms enable frictionless yet controlled user-to-application access, Netzilo AI Edge enables policy-driven, frictionless connectivity between AI agents and MCP servers. Access decisions can be dynamically enforced based on multiple contextual parameters, including agent identity, execution posture, tool usage, trust level of MCP servers, and real-time security signals from other protection layers.




Netzilo
Netzilo Inc.
166 Geary Str STE 1500 #2226
San Francisco, California, 94108
www.netzilo.com
sales@netzilo.com
+1 415 985 2636