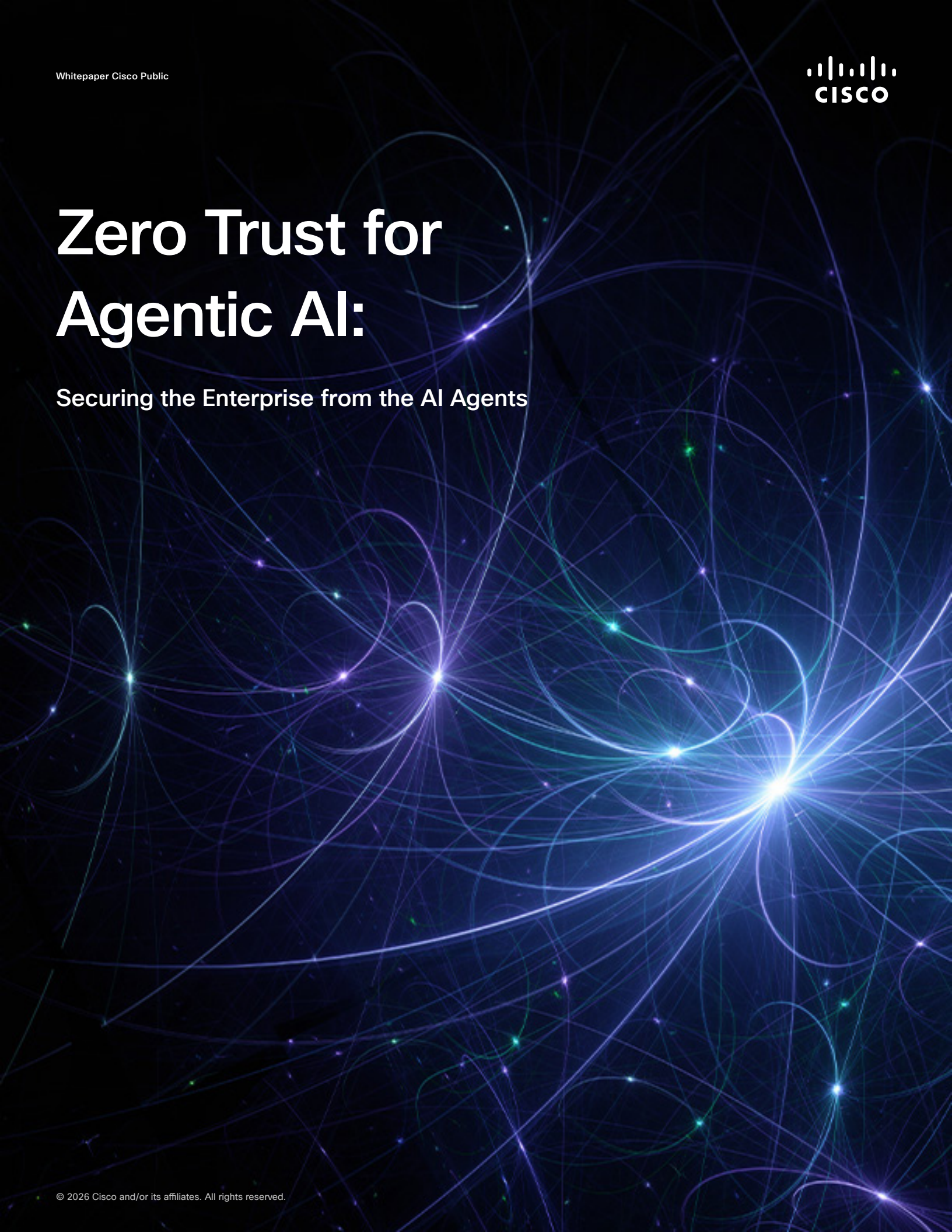
Zero Trust for Agentic AI:

Securing the Enterprise from the AI Agents



Executive Summary

Agentic AI is becoming embedded in enterprise operations. These systems reason, plan, and act autonomously across networks, executing workflows, accessing sensitive data, and interacting with tools at machine speed. As organizations move from experimentation to production, agentic AI is no longer theoretical. It is becoming a core operational capability across enterprise systems and workflows.

The challenge is not autonomy itself. Autonomous systems deliver real value when applied responsibly. The challenge is maintaining visibility, control, and accountability as autonomy scales. Agentic AI often operates with elevated privileges and broad access to systems and data. Without clear controls, speed outpaces oversight, expanding the attack surface beyond what traditional, human-centric security models can manage and creating new security risks across interconnected systems, tools, and data sources.

This risk concentrates across three dimensions: identity, access, and behavior. AI agents represent a rapidly growing class of non-human identities (NHIs) that must be discovered, governed, and mapped to accountable owners. Their access to tools, data, and applications must be tightly constrained to prevent permission sprawl and unintended exposure. And their behavior, operating at machine speed, must be continuously monitored and enforced in real time to remain aligned with intent, policy, and compliance.

Cisco's Zero Trust for Agentic AI extends Zero Trust principles to AI agents. The framework allows you to know every agent, authorize every action, and adapt to risk in real time across first-party agents, third-party platforms, and widely adopted AI applications. By integrating identity, security, and networking, Cisco applies consistent guardrails inline across agent interactions with enterprise systems, tools, and services as agentic AI scales across environments.

Organizations can scale agentic AI with confidence, unlocking its benefits while preserving trust, accountability, and control, ensuring every agent identity, access request, and action remains continuously verified.

Agentic AI in the Enterprise

Defining AI Agents: A New Class of Operational Risk

Agentic AI introduces a new class of operational risk for the enterprise. Unlike traditional AI applications or chat-based assistants, agentic systems reason, plan, and act autonomously across networks, applications, and workflows. They can invoke tools and execute actions on behalf of users or business processes without continuous human oversight. This autonomy enables efficiency and scale, but it also exposes gaps in security and governance models designed for static software and human-driven activity and introduces new risks as agents interact with enterprise systems, APIs, and external services at machine speed.

Depending on design and deployment, agents may operate with elevated privileges and broad access to sensitive data. That data, often drawn from multiple enterprise systems, APIs, and external services, provides the context that enables intelligent action. Without clear constraints on how agents access, use, and share it, organizations expand the attack surface beyond its intended boundaries, increasing the risk of unintended exposure or misuse across interconnected tools, platforms, and workflows.

Each agent operates as a Non-Human Identity (NHI), capable of authenticating, requesting resources, and interacting with tools and services. Unlike human users, these identities operate continuously, at scale, and at machine speed. Without effective discovery, visibility, and lifecycle governance, agents become unmanaged digital actors executing actions in the background with limited auditability and unclear ownership.

Traditional security models struggle in this environment. Controls built for static applications and human users assume predictable behavior, fixed access patterns, and manual oversight. Agentic systems break these assumptions. Agents dynamically call APIs, chain external tools, and interact with other agents. This creates distributed activity across agent frameworks, tool integrations, and emerging standards such as the Model Context Protocol (MCP), which conventional identity and access controls alone cannot effectively constrain.

As agentic AI is deployed across cloud, SaaS, and on-premises environments, these challenges intensify. Visibility gaps widen, privilege boundaries blur, and the risk of unintended or unsafe actions increases. Without consistent enforcement across identity, access, and behavior, a single agent misstep can propagate rapidly, leading to operational disruption, data exposure, or compliance violations that spread across interconnected systems at machine speed.

Addressing this shift requires a security model built for autonomous systems operating at machine speed. Zero Trust for Agentic AI applies continuous verification, precise access control, and real-time behavioral guardrails, restoring visibility, accountability, and control as agentic systems scale across the enterprise while ensuring agent identities, tool access, and runtime behavior remain continuously verified and governed.

Expanding Risk: The Growth and Complexity of Agentic Systems

Agentic AI has evolved from isolated assistants into distributed systems coordinating actions across cloud, SaaS, and on-premises environments. These agents combine large language models with plugins, APIs, and enterprise tools to form interconnected ecosystems that operate continuously and at scale. As autonomy increases, controlling where agents run, what they access, and how their actions propagate becomes significantly more complex as agent decisions trigger automated interactions across multiple systems and services.

Agents routinely interact with both internal enterprise systems and external services. They may access sensitive data such as customer records, intellectual property, or financial systems, while also connecting to SaaS platforms, partner APIs, and public data sources. Each integration expands the agent's footprint and introduces new trust assumptions. While this flexibility enables powerful workflows, it also multiplies the points where access can be misused, data can leak, or dependencies can be compromised across interconnected applications and data sources.

Risk compounds when agents coordinate through mechanisms such as the Model Context Protocol (MCP) or Agent-to-Agent (A2A) interactions. These frameworks enable context sharing, task delegation, and chained actions by allowing agents to invoke tools, retrieve context, and execute tasks across distributed systems. Still, they also introduce new control planes where permissions can drift, accountability can blur, and unintended or malicious instructions can propagate at machine speed. Without continuous visibility and enforcement, security teams lose the ability to trace decisions or intervene before they have an impact across these interconnected agent workflows.

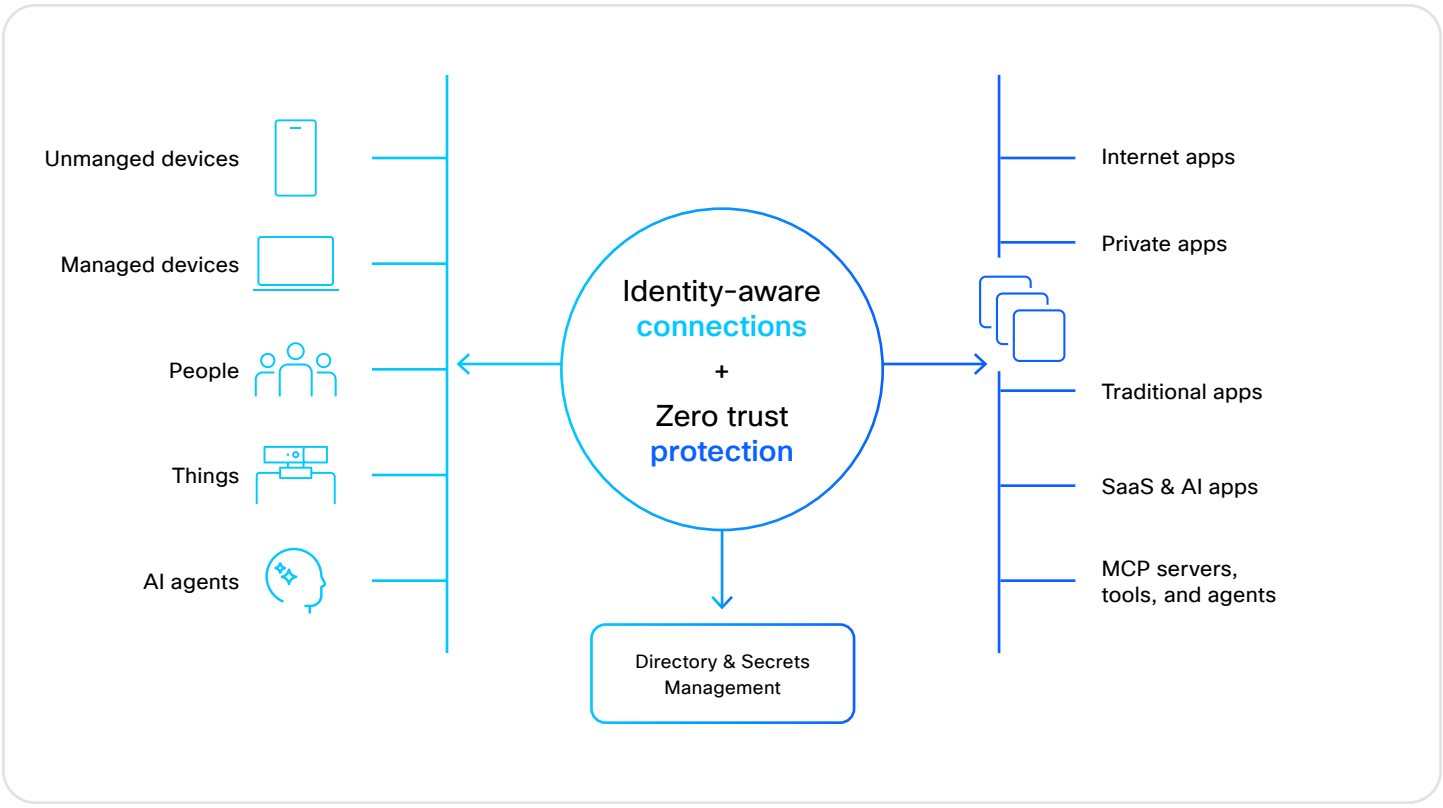
As agentic workflows scale across teams and business units, these interactions grow exponentially. A single misconfigured agent, compromised tool, or untrusted dependency can cascade across systems, expanding the blast radius of an incident well beyond its original scope. In this environment, fragmented controls become a liability. Identity, access, and monitoring applied in isolation cannot keep pace with distributed autonomy.

Managing this complexity requires security that operates at the same speed and with the same adaptability as the agents themselves. Enterprises need consistent visibility into agent identities, precise control over access, and real-time behavioral guardrails that evolve as interactions unfold. These requirements form the foundation of Cisco's Zero Trust for Agentic AI, enabling organizations to scale autonomy without losing control while ensuring every agent interaction, tool invocation, and system request is continuously verified and governed.

The Solution: Cisco Zero Trust for Agentic AI

To realize the full potential of agentic AI, organizations must ensure that every agent action is controlled, auditable, and aligned with business intent. As agents operate autonomously and at machine speed, security can no longer depend on static policies or manual oversight. Control must be continuous, enforced wherever agents interact with data, tools, and systems and applied inline as agent requests move across enterprise networks and services.

Cisco's Zero Trust for Agentic AI extends Zero Trust principles beyond people and devices to autonomous systems operating across the enterprise. The framework guides agents to perform only approved actions within defined security, privacy, and compliance boundaries by continuously verifying identity, enforcing precise access, and monitoring behavior in real time. In doing so, it closes the control gap introduced by autonomy and sees to it that agent actions remain continuously verified before they reach enterprise systems and tools.



Cisco's advantage lies in integration. By unifying networking, identity, and security, Cisco delivers end-to-end protection for agentic AI without relying on fragmented point controls. This approach applies consistently across enterprise-developed agents, agents hosted on third-party platforms, and widely used no-code applications such as ChatGPT, Gemini, Microsoft Copilot, and Perplexity AI. By enforcing Zero Trust policies across the network path where agent interactions occur, Cisco provides visibility and control over how agents access tools, APIs, and enterprise services.

Through Zero Trust for Agentic AI, enterprises can:

- **Discover and govern agent identities** across environments by maintaining a comprehensive inventory of AI agents, Model Context Protocol (MCP) servers, and connected tools and mapping each agent to an accountable human owner.
- **Enforce precise access controls** for Non-Human Identities (NHIs), applying least-privilege and time-bound access while validating the trustworthiness of models, tools, and dependencies before agents interact with enterprise resources. Apply access policies based on intent of agents as well.
- **Monitor and protect agent behavior** in real time using runtime guardrails on agent requests, responses, and tool invocations to detect and prevent unsafe, unauthorized, or anomalous actions.
- **Unify control across identity, network, and application layers**, eliminating blind spots as agents move between environments, tools, and workflows forconsistent enforcement regardless of where agents run.

This unified approach transforms AI security from a reactive safeguard into a proactive enabler. By embedding control directly into how agents operate, Cisco allows organizations to scale agentic AI with confidence, preserving visibility, accountability, and trust as autonomy accelerates while ensuring agent identities, access requests, and runtime behavior remain continuously verified.

Zero Trust for Agentic AI

Cisco's Zero Trust for Agentic AI is built on three interdependent pillars: Know every agent, Authorize every action, and Adapt to risk in real time. Together, they address the core risk dimensions of autonomous systems: identity, access, and behavior. Applied together, they provide the visibility, control, and guardrails required to secure agentic AI operating at machine speed. Applied in isolation, they leave gaps that erode trust and accountability.

These pillars operate continuously, not sequentially. Each reinforces the others, ensuring every agent identity is known, every interaction is authorized, and every action is monitored and enforced as autonomy scales across the enterprise.

1.Know every agent

Identity and oversight for every agent

Visibility is the foundation of Zero Trust for agentic AI. Organizations cannot govern what they cannot see, yet many lack a clear view of where AI agents are running, what resources they access, or who is accountable for their actions. Without this baseline, risk assessment is incomplete, compliance is harder to demonstrate, and trust cannot be established across increasingly distributed agent ecosystems.

Cisco addresses this gap by enabling continuous discovery, inventory, and monitoring of AI agents and their supporting Model Context Protocol (MCP) servers across hybrid environments, combined with strong identity assurance and device trust provided by Cisco Duo.

Each agent is treated as an NHI and mapped to a responsible human owner and provided appropriate access policies, establishing clear accountability and an auditable record of activity for every agent operating across enterprise systems and services.

This visibility provides the context required to enforce access controls and monitor behavior effectively. It is not simply an inventory function. It is the prerequisite for applying policy, managing risk, and sustaining trust as agentic systems evolve and interact with tools, APIs, and enterprise data sources.

Key Capabilities

- Comprehensive discovery and inventory of AI agents, models, and MCP servers, and connected tools across environments
- Human-to-agent ownership mapping to support accountability, auditability, and compliance
- Assign access policies to agents
- Continuous visibility into agent activity, data usage, and tool interactions across enterprise systems and services

2. Authorize every action

Applying least privilege and trust at every interaction

Once agents are discovered and identified, they must be governed with the same rigor applied to any enterprise identity. Each AI agent should have access only to the minimum required to perform its intended function. In agentic environments, unmanaged access becomes a primary risk, as over-privileged agents can move data, invoke tools, or modify systems at machine speed across interconnected enterprise systems and services.

This pillar governs the full lifecycle of NHIs, from creation and deployment through operation and retirement. Governance goes beyond static permission assignment. It requires continuous enforcement of fine-grained authorization policies that reflect agent intent, context, and scope, along with validation of the components agents depend on to operate.

Cisco enables least-privilege, intent-aware, and time-bound access controls that precisely constrain how agents interact with tools, data, and applications, delivered through Cisco Secure Access to enforce consistent, context-aware policy across users, agents, applications, and services. Access decisions are context-aware and enforced consistently across environments, reducing permission sprawl and unintended exposure. Cisco also validates the trustworthiness of MCP servers, third-party models, and integrated toolchains, thereby limiting the blast radius of misconfiguration or supply-chain compromise before agents are allowed to interact with enterprise resources.

By enforcing precise access control at every agent interaction, organizations have safeguards so autonomy does not become unchecked privilege, preserving control as agentic systems scale and evolve.

Key Capabilities

- Just in time access to agents so that they can perform their tasks in a timely manner
- Just enough access to agents to tools, resources and data
- Just long enough access to agents to perform their tasks

3. Adapt to risk in real time

Runtime guardrails for safe and compliant operations

Agentic AI operates autonomously and at machine speed, leaving little room for delayed detection or manual intervention. Even well-designed agents can behave in unintended ways due to ambiguous instructions, shifting context, or external manipulation. Behavioral protection addresses this risk by ensuring agent actions remain aligned with policy, intent, and compliance requirements through continuous runtime monitoring and enforcement.

This pillar focuses on evaluating what agents do as they act, not just who they are or what access they have. Cisco applies runtime guardrails, semantic inspection, and continuous monitoring to assess agent requests, responses, and tool invocations as they occur, delivered through Cisco AI Defense to provide runtime protection for agent interactions and workflows. By analyzing context and intent, these controls detect unsafe, unauthorized, or anomalous behavior before actions reach enterprise systems or propagate across agent workflows.

Real-time behavioral protection is essential for defending against threats unique to agentic AI, such as prompt injection, memory poisoning, and the misuse of legitimate tools. It also serves as a backstop for misconfiguration or policy gaps, preventing minor errors from escalating into broader incidents or expanding the blast radius across interconnected systems. These controls operate without disrupting legitimate workflows, preserving the speed and value of agentic systems while enforcing clear boundaries.

By embedding behavioral enforcement directly into runtime operations, organizations can contain risk as it emerges. This final layer of control completes the Zero Trust model for agentic AI by continuously verifying and governing agent behavior at machine speed.

Key Capabilities

- Real-time monitoring of intent, threat and behavior and enforcement of agent actions at execution time
- Semantic and contextual inspection of agent interactions to identify unsafe, unauthorized, or anomalous behavior
- Runtime protection with automated blocking or containment of high-risk activities before they impact enterprise systems

Technical Deep Dive: Securing Agentic AI Across the Lifecycle

Agentic AI is increasingly embedded in critical enterprise workflows, including customer support, cybersecurity operations, software development, and financial processes. In these domains, agents operate autonomously, process sensitive data, and interact with both enterprise and third-party systems through APIs, tools, and automated workflows. While this autonomy drives efficiency and scale, it also introduces new exposure across identity, access, and runtime behavior as agent actions propagate across interconnected systems at machine speed.

Cisco's Zero Trust for Agentic AI applies continuous control across the entire agent lifecycle. Instead of treating identity, access, and behavior as separate concerns, the framework governs who an agent is, what it can access, and how it behaves at machine speed across every interaction with enterprise systems, tools, and services. This lifecycle-based approach helps security keep pace with autonomy as agents evolve and scale across environments while maintaining consistent visibility, authorization, and runtime enforcement.

1. Know every agent: Securing Non-Human Identities

The Challenge:

Every AI agent operates as a NHI, capable of authenticating, requesting resources, and acting on behalf of the organization. Unlike human users, these identities operate continuously, scale rapidly, and exist entirely in software. Without formal identity management and visibility, agents become unmanaged digital actors with access to systems and data but no clear ownership or oversight.

Agents are often created dynamically as teams experiment with new workflows or integrate third-party tools. Without consistent discovery and lifecycle governance, organizations cannot reliably answer basic questions about agent existence, access, or accountability across increasingly distributed agent ecosystems. These gaps undermine risk assessment, auditability, and trust.

Key Risks

- **Identity lifecycle gaps:** Agents are often created without registration or decommissioning, leaving behind orphaned credentials, unused tokens, and unmanaged identities that continue interacting with enterprise systems.
- **Privilege escalation:** Static roles or over-permissive API keys grant agents broader access than required for their intended function across tools, APIs, and enterprise services.
- **Accountability gaps:** Without mapping agents to responsible human owners, organizations cannot trace actions back to an accountable party or establish clear audit trails for agent activity.
- **Credential theft:** Compromised NHI tokens or API keys provide attackers with legitimate, high-privilege access that is difficult to distinguish from normal agent activity and may allow malicious actors to impersonate trusted agents.

Zero Trust in Action

Cisco applies Zero Trust principles to agent identity through lifecycle management that mirrors human IAM processes. Visibility begins with continuous discovery and inventory of AI agents, Model Context Protocol (MCP) servers, and model integrations across environments. Each agent is registered as a first-class NHI and mapped to a responsible human owner with full lifecycle traceability and auditable ownership and assigned access policies.

Identity posture is continuously evaluated to detect anomalies such as credential reuse, unexpected access patterns, or orphaned identities across agent interactions with enterprise systems and services. When risk is identified, enforcement actions can be applied immediately to contain exposure before compromised credentials or unmanaged identities can be misused.

Example:

In a development environment, a build-automation agent that accesses private code repositories is registered as an NHI with narrowly scoped permissions. If its API key is reused or cloned outside of expected patterns, Cisco's identity analytics detects the anomaly through continuous monitoring of agent authentication and access behavior. It automatically suspends the agent's access, preventing further interaction until the issue is resolved.

Controls & Capabilities

- Continuous discovery and risk-scored inventory of AI agents, MCP servers, and model integrations across enterprise environments
- Centralized NHI directory with full ownership mapping and auditable identity history for every agent identity
- Assignment of access policies for agents
- Context-aware identity enforcement and ongoing privilege review to detect anomalies

2. Authorize every action: Managing Agent Access

The Challenge:

Securing agentic AI extends beyond runtime behavior monitoring. It requires governing how agents are created, what resources they are authorized to access, and how permissions change over time. Every agent operates within an ecosystem of models, plugins, APIs, and external services. Together, these dependencies form where weaknesses in any component can cascade across systems at machine speed as agents invoke tools and interact with distributed services.

In practice, access is often granted through static roles, long-lived credentials, or inherited development permissions. As agents move into production, access tends to expand rather than contract. Without continuous governance, organizations struggle to contain permission sprawl, validate dependencies, or detect agents operating outside their intended scope across enterprise systems and services.

Key Risks

- **Misconfigured environments:** Cloud services, APIs, or toolchains may grant agents broader access than required for their function, allowing unintended actions across enterprise resources.
- **Rogue or untracked agents:** Experimentation and rapid iteration can produce agents that bypass formal onboarding and governance processes and operate without proper identity registration or access controls.

Zero Trust in Action

Cisco applies Zero Trust principles to agent access control by enforcing least-privilege access, ensuring context-aware access, and continuously validating supply chain integrity. MCP servers, APIs, and integrated tools are assessed for trustworthiness before agents are allowed to interact with them. Agents that exceed authorized scope, violate policy, or rely on unverified components are flagged or blocked before actions reach enterprise systems or services.

Access policies are enforced consistently across environments and evolve with the agent lifecycle, reducing over-privilege as agents scale or change roles while maintaining continuous verification of agent interactions.

Example:

A finance automation agent is authorized to read data only from approved financial systems and generate reports to a validated destination. If it attempts to access an unsanctioned API or invoke an unapproved tool, Zero Trust policies block the request and log the event for investigation, preventing unintended data exposure or downstream impact across financial systems.

Controls & Capabilities

- Just in time access to agents
- Just enough access to agents
- Just long enough access to agents

3. Adapt to risk in real time: Enforcing Runtime Guardrails

The Challenge:

Agentic AI systems operate autonomously and at machine speed, often executing decisions without human review. This autonomy enables scale and efficiency but sharply reduces the time available for detection and response. Minor misconfigurations, ambiguous instructions, or malicious manipulation can escalate quickly, propagating across tools and systems as agents interact with distributed services and automated workflows.

Because agent behavior is non-deterministic and context-dependent, static rules and post-incident analysis fall short. Security controls must evaluate actions as they occur, analyzing intent, context and expected behavior in real time to prevent unsafe outcomes before they propagate.

Key Risks

- **Prompt injection and jailbreaks:** Hidden or crafted inputs that manipulate agent behavior or tool invocation to bypass safeguards or trigger unauthorized actions.
- **Memory poisoning:** Corrupted or altered context data that leads to persistent unsafe or unintended actions across future agent decisions.
- **Tool misuse:** Legitimate tools used in ways that exceed authorization, including data exfiltration or unauthorized system changes within enterprise systems or external services.
- **Agent-to-agent manipulation:** Compromised agents influencing other agents within multi-agent workflows, amplifying impact and expanding the blast radius across interconnected systems.

Zero Trust in Action

Cisco enforces behavioral protection through continuous runtime monitoring, semantic inspection, and automated response. Agent requests, decisions, and tool invocations are evaluated in real time against policy, context, and expected behavior to detect unsafe or unauthorized actions before they impact enterprise systems. When unsafe or anomalous activity is detected, actions can be blocked, privileges revoked, or the agent isolated to prevent further impact across connected tools and services.

These controls operate inline and at machine speed, allowing security to keep pace with autonomous execution while maintaining visibility and enforcement across agent interactions without disrupting legitimate operations.

Example:

A customer support agent processing untrusted external input encounters a prompt-injection attempt to alter its behavior. Cisco's semantic inspection detects manipulation in real time, prevents unsafe actions, and isolates the affected session, preserving service availability while protecting sensitive data from unauthorized access or exposure.

Controls & Capabilities

- Real-time inspection of agent requests, responses, and tool invocations to detect unsafe intent and policy violations
- Behavioral baselining to identify deviations from expected patterns of activity across agent workflows
- Automated blocking, containment, or isolation of unsafe or anomalous actions before they propagate across systems

Putting It All Together

Cisco's Zero Trust for Agentic AI unifies identity, access governance, and behavioral protection into a single, continuous framework. Every agent is discovered and accountable. Every interaction is authorized with the least privilege. Every action is evaluated and enforced in real time across tools, systems, and agent workflows.

By applying continuous verification across the agent lifecycle, this approach turns agentic AI from a potential liability into a trusted enterprise capability. Organizations can scale autonomy without losing visibility or control, enabling agentic systems that are observable, accountable, and secure by design.

Deployment Scenarios: Applying Zero Trust Across Every Agentic Environment

Agentic AI spans multiple enterprise environments, from custom-built agents on internal infrastructure to third-party and no-code applications adopted by business teams. While each model introduces distinct risks, the control requirements are consistent. Cisco's Zero Trust for Agentic AI applies the same identity, access, and behavioral guardrails across all environments, enabling secure adoption without fragmented visibility or inconsistent policy enforcement.

1. First-Party Agents: Enterprise-Developed and Hosted

Description

First-party agents are developed by the organization and deployed within its own cloud or on-premises environments. They automate internal workflows such as customer support, IT operations, and DevSecOps, often integrating directly with proprietary systems and sensitive data through APIs, enterprise tools, and automated workflows.

Security Focus

Because the enterprise controls both the agent and its infrastructure, the primary challenge is governance. Organizations must ensure that only sanctioned agents operate, that access is tightly scoped, and that every action is attributable to an accountable owner through continuous identity verification and activity monitoring.

Key Controls

- **Agent and MCP server discovery:** Continuous inventory of all agents, models, and MCP endpoints across enterprise environments.
- **Identity governance:** Each agent registered as a NHI with human ownership traceability and auditable lifecycle management.
- **Runtime guardrails:** Safety and privacy protections that prevent misuse of data or unsafe actions through continuous monitoring and enforcement of agent behaviors.

Example

An internal IT operations agent automates system patching by interacting with network configuration tools. Cisco's Zero Trust controls guides the agent to operate only within approved boundaries, log all activity for auditability, and is prevented from accessing unrelated systems or data through continuous verification of its identity, access permissions, and runtime behavior.

2. Platform-Hosted Agents: Developed on Third-Party Platforms

Description

Many enterprises build or extend agents on cloud-based platforms such as AWS AgentCore or Azure AI Studio. These agents operate within third-party infrastructure while still interacting with internal enterprise systems, APIs, and sensitive data across organizational trust boundaries.

Security Focus

The shared-responsibility model introduces both integration and supply chain risk. Organizations must validate the trustworthiness of third-party components, maintain visibility into agent behavior, and enforce strict privilege boundaries as agents interact with internal resources and external services.

Key Controls

- **Identity and access management:** Context-aware authentication and least-privilege permissions for agents operating on cloud platforms and interacting with enterprise systems.
- **Supply chain validation:** Continuous risk assessment of hosted MCP servers, APIs, and third-party plugins to verify trusted components before agent execution.
- **Agent-to-human mapping:** Clear accountability by linking each platform-hosted agent to a verified enterprise owner and registering the agent as a NHI.

Example

A finance team builds a cost-optimization agent using AWS AgentCore. Cisco's Zero Trust controls validate the associated MCP server, enforce least-privilege API access, and continuously monitor the agent's actions as it interacts with internal financial systems and cloud services to prevent data leakage or unauthorized spending.

3. No-Code and Third-Party Agentic Applications

Description

Employees increasingly adopt no-code AI tools such as ChatGPT, Gemini, Microsoft Copilot, or Perplexity AI to improve productivity. These agents may access enterprise information through user inputs, file uploads, or API integrations, often outside centralized IT oversight and traditional security controls through browser-based and external AI services.

Security Focus

This deployment model introduces the highest level of shadow AI risk. Usage frequently occurs beyond formal governance frameworks, increasing the likelihood of uncontrolled data exposure through prompts, file uploads, or unsanctioned integrations. The primary priorities are discovery, access enforcement, and runtime monitoring to ensure sensitive enterprise information remains protected.

Key Controls

- **Agent discovery and visibility:** Identify sanctioned and unsanctioned no-code agents in use across the organization and detect emerging AI application usage.
- **User identity management:** Enforce secure authentication and data-handling policies for users interacting with external AI tools through enterprise identity controls.
- **Behavioral guardrails:** Monitor and restrict agent interactions with corporate systems and data sources to maintain compliance and prevent unauthorized sharing of sensitive information.

Example

An employee uses a generative AI assistant to summarize customer feedback from internal reports. Cisco's secure web gateway detects and controls interactions, ensuring sensitive data is not transmitted to unapproved systems or external platforms while allowing legitimate productivity use cases to continue.

Unified Protection Across All Deployments

Across all deployment models, Cisco's Zero Trust for Agentic AI applies the same continuous verification approach for secure, governed operation at scale across enterprise systems, tools, and agent workflows. Whether agents are enterprise-developed, platform-hosted, or user-adopted, the framework enforces consistent controls:

- Every agent is discovered and identifiable as a governed NHI
- Every connection is authenticated and authorized with context-aware, least-privilege access control
- Every action is monitored and governed in real time through continuous behavioral inspection and enforcement

This unified coverage enables organizations to scale agentic AI safely, accelerating innovation while maintaining confidence in the security, visibility, and integrity of their AI ecosystem.

Why Cisco

Cisco is uniquely positioned to secure the era of agentic AI.

Agentic systems demand security that spans identity, access, behavior, and the network. Cisco uniquely brings these capabilities together into a unified Zero Trust framework that secures both human users and autonomous AI agents across the enterprise. This integrated approach delivers end-to-end protection, from agent discovery and identity governance to runtime behavioral enforcement, across networks, applications, and cloud environments.

Built on decades of leadership in Zero Trust, secure access, and AI-driven security, Cisco enables continuous verification at machine speed. Organizations gain a consistent layer of trust across networks, clouds, and AI systems, allowing them to scale agentic AI securely, responsibly, and with confidence while maintaining visibility and control across distributed AI ecosystems.

For more information, contact your Cisco representative or visit: https://cisco.com/go/securing-agentic-ai.