

Protect Government Agentic AI Against Nation-State Attacks

How AgentMinder Helps Agencies Govern Agent Identity, Runtime Actions, and Mission Boundaries

Thank you for your interest
in exploring this content.

Carahsoft is the **Trusted Government IT Solutions Provider**® supporting a broad portfolio of industry-leading technologies through OH MSCA, NASPO ValuePoint, Texas DIR-CPO-5683 and a wide range of other contract vehicles.

As the **Master Government Aggregator**®, Carahsoft connects government agencies, industry partners, and technology providers to deliver innovative, mission-focused solutions.

In partnership with Broadcom, we provide technology solutions that drive modernization, strengthen operations, and ensure compliance with evolving government standards.



To learn more about how Carahsoft can support your technology needs, please visit carahsoft.com



Explore More Resources:
carah.io/BroadcomResources



Join Events & Webinars:
carah.io/BroadcomEvents



Discover Technology Solutions:
carah.io/BroadcomSolutions



Learn About Procurement:
carah.io/BroadcomContracts



Connect With Our Team:
Broadcom@carahsoft.com
(571) 662-3260

Solution Brief

Challenges

Autonomous agents expand the government attack surface across mission systems, sensitive data, delegated authority, and connected tools.

Opportunities

Governed agents can accelerate public services and mission workflows without granting broad, persistent access to critical systems.

Business Outcomes

Runtime controls and audit evidence empower agencies to scale agentic AI, drive mission efficiency, reduce operational risk, and safeguard sensitive data.

Protect Government Agentic AI Against Nation-State Attacks

How AgentMinder Helps Agencies Govern Agent Identity, Runtime Actions, and Mission Boundaries

Overview

Government agencies are beginning to deploy agentic AI to accelerate public services, mission workflows, and critical infrastructure operations. Unlike traditional assistants, these systems can retrieve records, invoke tools, update systems, and trigger downstream decisions with limited human oversight. That shifts the security problem from governing model output to governing machine-executed authority.

For public-sector organizations, the consequences of compromise are significant. A manipulated or stolen agent identity can expose sensitive data, misuse tools, disrupt operations, or undermine trust in agency decisions. Recent NIST and CISA guidance makes clear that secure deployment of AI agents requires stronger controls for identity, authorization, monitoring, and runtime governance, especially where public services and critical infrastructure are involved.

AgentMinder helps agencies address that challenge by providing a dedicated control plane for agent identity and authorization. It gives every agent a verifiable identity, enforces mission-scoped and intent-aware authorization on every tool call, and preserves a full audit trail for investigation and accountability. The result is a more governable foundation for deploying agentic AI in sensitive government environments.

Attack Paths

Nation-state actors do not need to compromise the model itself to compromise an agentic AI system. In practice, they are more likely to target the agent's inputs, credentials, tool connections, or delegated trust relationships. For government agencies, that can mean unauthorized access to sensitive data, misuse of public-service workflows, disruption of mission operations, and loss of accountability in high-consequence environments.

- **Prompt injection and tool misuse:** Attackers can hide instructions in prompts, documents, emails, or retrieved content that steer an agent into invoking tools, exposing data, or taking unauthorized actions.
- **Credential theft and replay:** If an agent's token, certificate, or client credential is stolen, an attacker may be able to impersonate that agent and move through trusted workflows unless identity is tightly bound and narrowly scoped.
- **Delegation abuse and lateral movement:** In multi-agent workflows, a compromised agent can become a pivot point for broader compromise if delegation is not explicit, constrained, and attributable.
- **Tool impersonation:** A fake or spoofed MCP server, API, or service endpoint can intercept requests, manipulate results, or capture credentials if agents are allowed to connect without strong verification.
- **Supply-chain compromise:** Trusted tools, plugins, or data sources can be poisoned after approval, allowing malicious behavior to spread across multiple agents and workflows.

AgentMinder Addresses Agentic AI Attack Vectors

Attack Vector	AgentMinder Response
Prompt Injection/Tool Misuse	The AI gateway enforces policy independently of the LLM, so a manipulated prompt cannot directly invoke unauthorized tools or actions.
Credential Theft/Replay	Mission-scoped, time-bound sessions and cryptographically bound agent identity reduce replay and limit what a stolen token can do.
Lateral Movement	Delegation chains are explicit and constrained, so child agents inherit narrower authority and cannot silently escalate privileges.
Tool Impersonation	Agent identity, certificate-based trust, and gateway-controlled routing help ensure agents connect only to approved back ends and services.
Supply-Chain Compromise	Per-call policy evaluation, runtime telemetry, pre-approved tool reachability, intent checks, and anomaly detection help identify abnormal behavior and stop unsafe actions during a mission.

In each case, the architectural question is the same: *Can the agency verify who the agent is, what mission it is operating under, which tools it is allowed to reach, and whether each action is still authorized at runtime?*

How the Control Plane Responds

Agentic AI security cannot rely on prompts, static permissions, or post-incident review alone. In government environments, agencies need a control plane that continuously verifies agent identity, constrains authority to an approved mission, enforces policy on every tool call, and preserves the evidence needed to investigate and revoke access in real time. This is the layer that separates an autonomous system from an ungoverned one.

- **Identity, mission, and per-call authorization:** AgentMinder evaluates every tool call against the agent's identity, declared mission, intended action, target resource, and policy context before granting access. Missions are time bound and revocable, and cryptographically bound credentials help reduce replay and impersonation risk.
- **Delegation, attribution, and identity propagation:** In multi-agent workflows, authority must remain attributable as work is delegated. AgentMinder constrains delegation depth, preserves lineage across parent and child agents, and allows administrators to control whether a delegated agent acts under its own identity or on behalf of a human principal, reducing the risk of silent impersonation and privilege expansion.
- **Tool governance and reconnaissance resistance:** Agents should not be able to discover or invoke every tool in the environment. AgentMinder limits visible tools to those the mission is authorized to use, verifies trusted connections, and can apply payload-level constraints to restrict sensitive fields, classified data, or out-of-scope requests. This helps reduce both reconnaissance and data-exfiltration risk.
- **Revocation and federated response:** Authorization must remain valid throughout the mission, not only at login. AgentMinder continuously checks that a mission remains active and fails closed when the control plane cannot confirm it. Through standards-based event sharing, agencies can also propagate revocation and compromise signals across federated environments, helping disable compromised agents before they continue operating elsewhere.
- **Observability and anomaly detection:** Runtime decisions are correlated with telemetry, audit events, and behavioral risk signals so security teams can detect abnormal activity, reconstruct agent behavior, and respond with stronger evidence during investigations or oversight reviews.

Why This Matters for Agencies

As agencies move from pilots to production with agentic AI, the question is no longer whether AI can improve efficiency, but whether it can do so without undermining core missions. When agents can access citizen records, trigger benefits, update case files, or interact with operational systems, a single compromise can lead to delayed services, unauthorized disclosure of sensitive data, disruption of critical infrastructure, or decisions that cannot be reliably traced back to a responsible party.

For public-sector leaders, this is a trust and accountability problem. Agencies need confidence that AI will not silently alter constituent outcomes, create new compliance exposures, or erode trust in digital services. Security and risk teams need a way to answer hard questions under scrutiny: which agent acted, under whose authority, using which tools, and against which data.

That requires more than application-level guardrails or static model policies. It requires a dedicated control plane for agents; one that assigns each agent a distinct identity, binds it to a defined mission, enforces what it may do at runtime, and produces a defensible audit trail for every action.

Why Broadcom

Broadcom brings more than a point solution to agentic AI security. AgentMinder is built as a dedicated control plane for agent identity, authorization, delegation, and audit, using a standards-based architecture that aligns with emerging models for secure AI agents. Support for interoperable approaches such as AuthZEN, DPoP, token exchange, SCIM, and CAEP helps agencies reduce proprietary lock-in, align with evolving government requirements, and extend security controls across federated environments.

Broadcom also brings the deployment flexibility government environments require. AgentMinder can run alongside AI models in government clouds, private clouds, or on-premises environments, helping agencies keep policy enforcement close to the workloads, data, and trust boundaries they already govern. For organizations operating shared infrastructure, isolated tenant boundaries and standards-based integration provide a stronger foundation for scaling agentic AI securely across departments, missions, and security domains.

Together, these capabilities give agencies a more complete foundation for governing autonomous agents across mission environments.

AgentMinder Differs from Conventional AI Security and IAM Approaches

Differentiator	What It Means	Why It Matters for Agencies
Intent-Aware Authorization	AgentMinder evaluates whether an individual tool call aligns with the agent's declared mission, identity, target resource, and policy—not merely whether a credential or API scope is valid.	Conventional IAM approaches often validate identity, credentials, roles, or API scopes. They may not evaluate whether a specific agent action remains aligned to its declared mission and policy context at runtime.
Mission-Scoped Delegation	Parent-child agent relationships retain bounded authority, so delegated agents cannot silently exceed the mission or authority of the originating agent.	Many conventional IAM and network-security models treat agents as flat service principals or API clients, limiting their ability to preserve and enforce inherited authority across parent-child agent workflows.
Runtime Enforcement on Every Tool Call	Policy is evaluated in the execution path, preventing a manipulated prompt or valid-but-misused credential from directly becoming an unauthorized action.	This provides active runtime governance beyond passive discovery, static posture management, WAF/rate limiting, or model scanning.
Private, Air-Gapped Deployment	AgentMinder can run alongside workloads in on-premises, private-cloud, VPC, and air-gapped environments without a public-SaaS control-plane chokepoint.	Many cloud-delivered control planes depend on external connectivity, creating challenges for disconnected, air-gapped, or tightly controlled private environments.
Forensic-Grade Agent Accountability	Correlated OpenTelemetry-based audit trails tie identity, mission, policy decision, tool call, and outcome together.	Standard network, posture, or IAM logs may not establish the mission context, policy decision, and authorization chain behind an autonomous agent action.
Open, Interoperable Control Plane	A pluggable policy-decision model using AuthZEN lets agencies retain existing IAM investments while adding agent-specific runtime governance.	Established enterprise IAM platforms remain important for workforce identity, directory services, authentication, and access governance. AgentMinder complements those investments by adding a dedicated runtime control plane for autonomous AI agents, where identity alone is not sufficient to govern mission context, delegated authority, and tool actions.