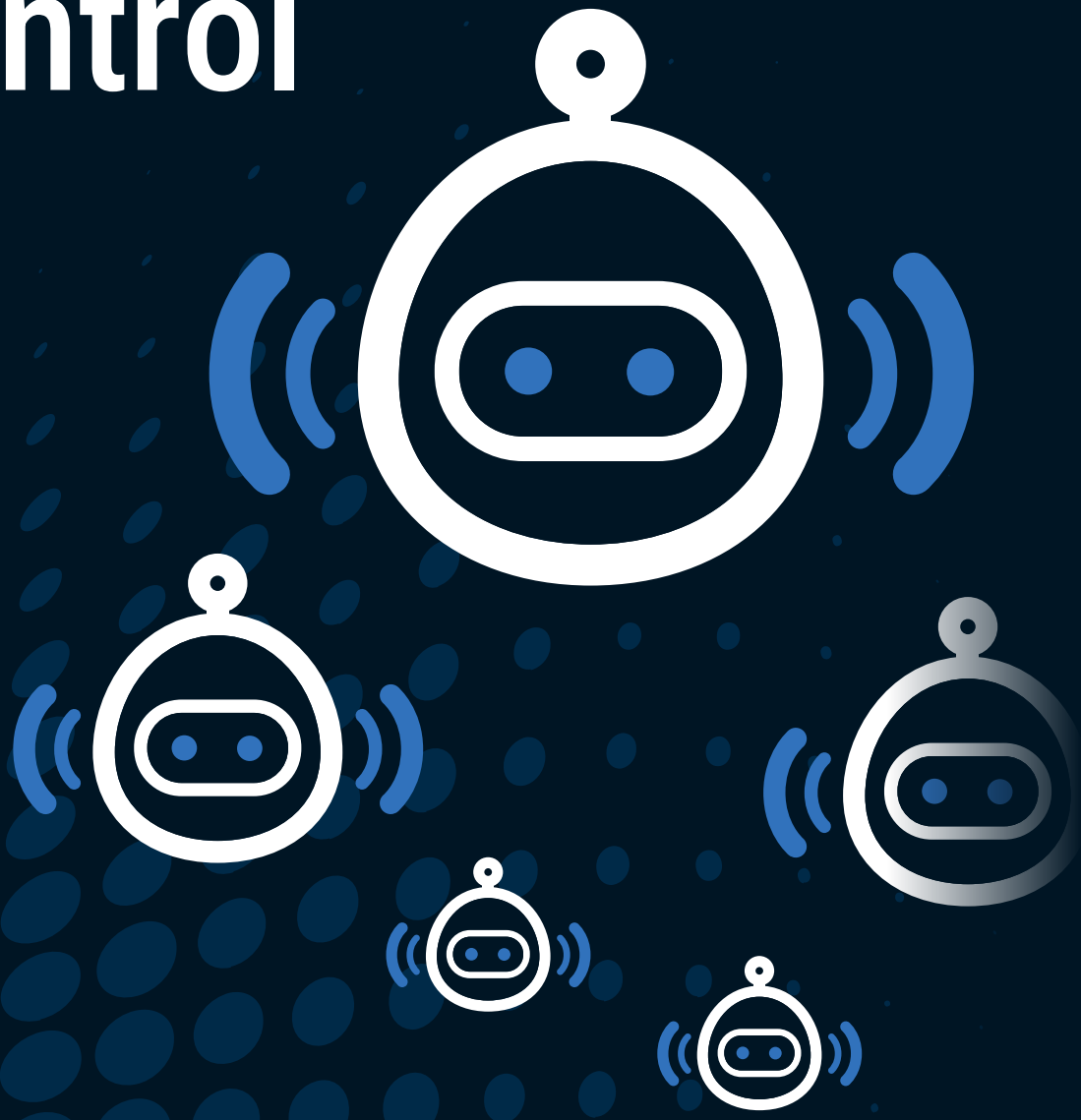
From AI Sprawl to Trusted Control

digicert®

Practical use cases for securing the expanding agentic surface

eBook | 2026



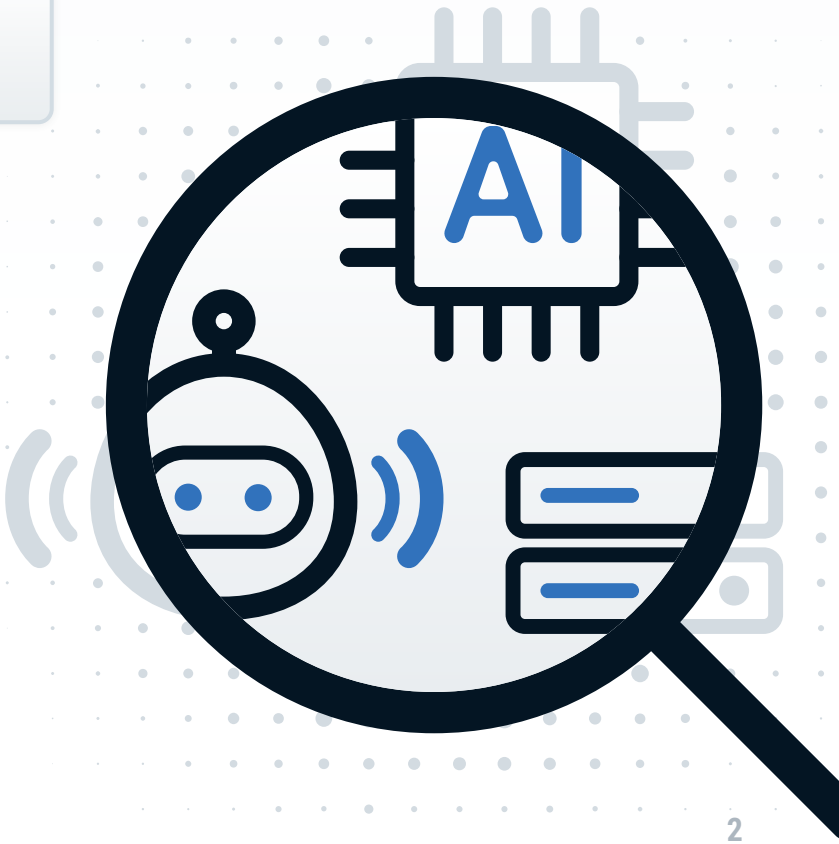
Introduction

AI is rapidly changing how organizations do business. Agents can interact with models, applications, APIs, data, tools, and even other agents, often with increasing levels of autonomy.

Meanwhile, models make business-critical decisions based on sensitive data, frequently with limited transparency.

As these AI assets proliferate, most enterprises lack a consistent way to answer fundamental questions:

- **What are the agents operating in my environment?**
- **What are they allowed to do?**
- **What happens when they behave outside of policy?**
- **Are the models we're relying on trustworthy?**



Extending trust to the new agentic surface requires a trust layer designed for AI systems, one that allows visibility and control over every agent, MCP server, and model across an organization's digital environments.

DigiCert® AI Trust provides that trust layer for enterprise AI assets.

Organizations can discover agents across their ecosystem, bring them into a centralized inventory, issue cryptographically verifiable AI Agent Passports, establish policy controls, monitor violations, revoke trust when an agent presents unacceptable risk, and ensure the integrity of key AI models.

Here are four real-world scenarios showing how organizations can move from guessing what AI assets exist to establishing and continuously governing trusted AI behavior.

Use case 1:

Enterprise AI discovery: Global retailer stops agent sprawl

Scenario

A global multi-brand retailer has multiple business units experimenting with generative AI.

Various teams are deploying agents across cloud environments to automate customer service, purchasing, merchandising, order management, software development, and operational support.

Security leadership asks a simple question:

How many AI agents are actually operating across the enterprise?

No team can provide a complete answer.

Challenge

AI agents can be created rapidly and deployed independently by application and cloud teams.

Without centralized visibility, security teams may not know:

- Which agents exist
- Where they are running
- What models and tools they use
- What data or resources they can access
- Which agents have established trusted identities
- Which agents are managed versus unmanaged

Many of these agents can access or even change production environments. They may have access to sensitive customer or payment data. They could create openings for breaches or simply act in ways detrimental to the business while in pursuit of their assigned tasks. In short, these agents create an expanding AI visibility and governance gap that can have serious consequences.

Why traditional approaches fall short

- Cloud asset inventories identify infrastructure, not necessarily the AI agents operating on top of it.
- Application inventories depend on teams registering applications manually and maintaining a regular inventory refresh schedule.
- Cloud IAM can identify workload credentials but does not provide a comprehensive view of an agent's identity, resources, purpose, and trust posture.

How AI Trust works

DigiCert AI Trust discovers AI agents within the enterprise ecosystem and brings them into a centralized agent inventory.

Security and platform teams gain a comprehensive view of the agent estate, including information about associated:

- Models
- Tools
- Guardrails
- Data resources
- Workloads and environments
- Trust and passport status

Discovered agents can then be brought under management, issued passports, and prepared for trusted identity and policy enforcement.

Business impact

- **Creates a centralized AI agent inventory for various agent types**
- **Establishes the foundation for enterprise AI governance**
- **Gives security teams visibility before enforcing control**
- **Exposes shadow AI and unmanaged agents**
- **Reduces AI blind spots**

Key takeaway

You cannot govern AI agents you cannot see. Trust begins with discovering the complete AI estate.

Use case 2:

Trusted agent identity: Financial institution sets agent guardrails

Scenario

A financial services company deploys a customer-facing AI agent that helps resolve customer service requests. To complete its work, the agent will interact with internal APIs, customer account systems, knowledge bases, and approved external services. The agent will be able to access sensitive customer and financial data.

The organization needs to know:

How can we prove this is the approved agent before granting access to systems and data, and how can we control exactly what it is allowed to access?

Challenge

NA traditional workload credential may authenticate the infrastructure running the agent, but it does not necessarily communicate the complete trust context of the agent itself.

This agent needs to interact with sensitive data, and the financial institution must be able to prove how the agent has been secured. Security teams need a way to establish a trusted agent identity and associate that identity with controls governing how the agent operates.

Why traditional approaches fall short

- API keys are shared secrets.
- Cloud IAM identities primarily represent infrastructure.
- Network controls can restrict connectivity, but they often lack agent-specific identity and business context.
- Application-specific access controls can become fragmented across every system the agent touches.
- Persistent access creates significant risks, especially when the agent is interacting with sensitive data and financial systems.

How AI Trust works

The organization brings the agent under management and issues a DigiCert AI Agent Passport. The passport provides the agent with a cryptographically verifiable, trusted identity. Reusable passport templates allow the organization to standardize both broad and granular trust controls.

Coarse-grained controls can govern areas such as:

- mTLS requirements
- Ingress and egress
- Rate limits
- Agent capabilities
- Operational constraints

Fine-grained Rego-based policies can then govern more specific authorization and behavioral requirements.

The result is a trusted control point that connects agent identity with policy.

Business impact

- **Supports regulatory compliance with clear audit trails tied to each individual agent**
- **Creates reusable governance patterns for different classes of agents**
- **Reduces reliance on fragmented application controls**
- **Enables consistent security controls across agents**
- **Supports least-privilege agent access**
- **Establishes verifiable agent identity**

Key takeaway

An AI agent should not simply have unrestricted access. It should be able to prove its identity and operate within explicitly defined trust boundaries.

Use case 3:

Risk response: Technology company neutralizes rogue agent

Scenario

An AI agent operating within a technology company's environment is repeatedly attempting actions that violate established policy.

The agent is generating policy denials against a restricted destination and behavior is inconsistent with its expected operating profile. This has every marker of being the next headline about rogue agents behaving badly.

The security team needs to quickly determine:

What happened, how serious is it, and do we have a way of stopping the agent before harm occurs?

Challenge

AI agents can operate continuously and at machine speed.

When an agent behaves unexpectedly, an alert sent to a security team may already be too late to stop a serious incident from occurring. Security teams need context about the agent, the policy being violated, where the agent was attempting to connect, and how frequently the behavior is occurring.

They also need a mechanism to rapidly remove trust when the risk exceeds organizational tolerance. Crucially, they need a way to do this before harm has occurred, so that an investigation may follow after the agent has already been neutralized.

Why traditional approaches fall short

- Traditional security alerts may show an infrastructure or network event without identifying the AI agent responsible.
- Alerts may also come too late, long after the rogue agent has already caused harm.
- Disconnecting infrastructure can impact unrelated workloads.
- Rotating credentials can take time and may not address the broader trust relationship.
- Investigations become fragmented across IAM, cloud, network, application, and logging systems, and often happen after a security incident has already occurred.

How AI Trust works

DigiCert AI Trust centralizes policy violations for managed AI agents and automates the kill switch. When an agent attempts an action outside of established policy, the passport is immediately revoked, sending the agent to quarantine and before harm occurs.

Security teams can then investigate violations after the fact using context such as:

- Affected agent
- Triggered policy
- Severity
- Destination
- Enforcement point
- Occurrence and timing information
- Agent and passport context

Based on the organization's risk profile, teams can remediate the underlying policy, investigate the agent further, or permanently revoke the Agent Passport. The kill switch triggers automatically based on defined policy to quarantine the agent and remove its trusted operating status. It can also be triggered manually at any time.

Business impact

- **Stops rogue agent behavior before harm occurs**
- **Centralizes AI agent risk visibility**
- **Enables risk-based remediation**
- **Provides rapid containment for compromised or misbehaving agents**
- **Connects policy violations directly to agent identity**
- **Accelerates investigation**

Key takeaway

Trust must be continuously earned. When an AI agent exceeds your risk tolerance, organizations need the ability to revoke that trust immediately.

Use case 4:

Model integrity: Healthcare organization protects diagnostics

Scenario

A healthcare technology organization has developed a specialized diagnostic model that processes imaging results to assist in earlier, more accurate diagnoses of complex, life-threatening medical conditions.

This model will be delivered to healthcare providers who will utilize it in clinical settings with real patient data.

To ensure better health outcomes, patient privacy, and intellectual property protection, the team must understand:

How can we prove this model remains trustworthy, and who can access and view the model data as well as the patient information it is processing?

Challenge

AI models introduce many new risks, especially in regulated industries such as healthcare.

Models may be at risk of tampering, unauthorized modification, drift, and integrity degradation. Proprietary models are at risk of intellectual property theft and redistribution. Data leakage may occur in insecure execution environments, exposing sensitive data.

These risks may impact model makers, model users, or both. AI models create a trust gap across the ecosystem that is difficult to bridge.

Why traditional approaches fall short

- Running the model in the cloud creates potential risks around data access. Security approaches may protect data in transit and at rest, but not while it is in use.
- Handing the model over to be run in the model user's private environment may solve for data privacy, but that creates the risk of unintended drift or even tampering which can impact outcomes. It also poses a risk of intellectual property theft for the model maker.
- Regardless of where the model is running, ensuring consistent outcomes requires proof that the algorithm running in deployment is the exact one that was initially deployed.

How AI Trust works

DigiCert AI Trust offers runtime protection for models to ensure integrity and confidentiality. Models run in trusted execution environments (TEEs), ensuring they remain encrypted in memory, stay isolated from the host operating system, and protect models and the data they use for inference.

The keys to ensuring provable model integrity are:

- Model provenance data
- Cryptographic encryption
- Secure key management
- Continuous model validation
- Version hashing and lineage tracking

All of this, combined with a confidential enclave for execution and lifecycle controls, can ensure that the model continues functioning as intended, and all relevant data is protected.

Business impact

- **Protects models from exfiltration, reverse-engineering, and redistribution**
- **Ensures execution occurs only in verified, trusted environments**
- **Prevents unauthorized access to sensitive data**
- **Protects data in transit, at rest, and in use**
- **Continuously validates runtime integrity**
- **Supports regulatory compliance**

Key takeaway

AI models must be protected at runtime and continuously validated to ensure accurate outcomes and protect all parties in the model ecosystem.

Key questions for trusted AI

As organizations across every industry move into production with AI agents and models, they must be able to answer **five fundamental questions**:

Question #1

What AI is running across all environments?

Question #2

What data flows into and out of AI assets?

Question #3

Who owns these assets, and what credentials or privileges do they hold?

Question #4

Can the organization prevent unauthorized behavior and revoke access if trust conditions change?

Question #5

Can actions be reconstructed to prove what happened?

These questions point to gaps in visibility, identity controls, auditability, and data controls.

Closing these gaps is imperative as organizations move into production with AI agents and models.



Accelerating trusted AI with DigiCert

DigiCert AI Trust Manager can help security and IT teams answer key questions about their agentic surface by delivering a single control plane for AI agents, models, and MCP servers.

AI Trust Manager simplifies discovery, federated identity, and lifecycle management for AI assets by providing:

- **Discovery and continuously refreshed inventory of AI assets**
- **Identity and policy bound into a single verifiable, revocable credential**
- **Control and enforcement mechanisms, including an automated kill switch**

To protect AI assets at internet scale, trust fabric for AI is powered by the DigiCert AI Root of Trust—a root that is already ubiquitous, now repurposed for AI.

Conclusion

As AI gains autonomy, trust must become something organizations can establish, verify, enforce, and revoke. Security teams need technical controls that expose what AI assets exist, define what they are trusted to do, verify trust continually, and take action when trust conditions change.

DigiCert AI Trust Manager gives organizations the tools they need to move from fragmented AI oversight to a consistent, verifiable trust model across agents, models, and MCP servers.

As AI becomes more autonomous, trust cannot simply be assumed. It must be established, continuously verified, and enforced.

Want to see DigiCert AI Trust Manager in action? Explore the interactive demo.

