

Why governing system-to-system trust will become a foundational layer of security architecture

AI Agents are Rewiring the Enterprise Architecture

AI Agents Are Rewiring the Enterprise

Why governing system-to-system trust will become a foundational layer of security architecture

Bob Moul and Mark Viglione

March 10, 2026

Bob Moul and Mark Viglione are cybersecurity entrepreneurs and the creators of the ZTNX (Zero Trust for Internal Networks) category.

EXECUTIVE SUMMARY

Artificial intelligence is rapidly becoming embedded in enterprise operations. AI copilots, autonomous agents, and retrieval-augmented systems are beginning to automate workflows that span multiple applications, APIs, and internal services. In doing so, these systems are fundamentally changing how enterprise infrastructure operates.

Historically, most enterprise activity followed predictable patterns: users accessed applications, applications accessed databases, and system relationships were relatively stable. Security architecture evolved to govern these interactions. Identity and Access Management (IAM) hardened authentication and authorization, while segmentation and detection tools helped limit and monitor network activity.

AI agents introduce a different model. Rather than simply accessing systems on behalf of a user, they autonomously orchestrate workflows across multiple platforms, APIs, and internal services.

Much of the current conversation around AI security focuses on model risks such as prompt injection, hallucinations, and data leakage. While these are important concerns, they overlook a broader architectural shift. As AI agents begin orchestrating workflows across enterprise systems, they dramatically increase the number of system-to-system trust relationships inside the network - often between systems that historically rarely communicated.

**As these relationships multiply, the enterprise trust surface expands accordingly.
Yet no security architecture was designed to continuously govern that trust.**

This creates a structural gap in enterprise security architecture.

Existing controls remain essential, but they address different questions. IAM governs authentication and permissions. Segmentation enforces predefined policies. Detection platforms identify suspicious behavior. But none are designed to continuously evaluate whether internal system relationships themselves are appropriate, necessary, and safe.

At the same time, many security teams are beginning to focus on identifying AI activity within their environments - attempting to detect RAG pipelines, agent orchestration traffic, and other AI-driven workflows. This visibility is valuable. However, it does not address the broader issue created by AI-driven automation.

**The goal should not simply be detecting AI traffic.
It should be governing the trust relationships AI systems create inside the enterprise.**

As AI adoption accelerates, the security challenge is shifting from simply authenticating identities to governing **system-to-system trust** across the enterprise. Organizations will increasingly require the ability to continuously validate which internal communications are necessary, appropriate, and safe.

In the same way that IAM established discipline around user access, a new layer of security architecture is emerging to govern internal system relationships. This evolution represents the next step in the Zero Trust journey - extending trust verification beyond identities to the communications that connect modern digital infrastructure.

**IAM governs trust for users and agents.
The next challenge is governing trust between the systems they connect.**

1. AI Is Changing How Enterprise Systems Interact

AI capabilities are rapidly moving beyond conversational assistants into systems that can autonomously perform tasks across enterprise environments. These AI agents can retrieve information, interact with APIs, trigger workflows, and coordinate activities across multiple applications and services.

Examples of emerging enterprise use cases include:

- AI assistants retrieving information from internal knowledge systems
- Automated support agents interacting with CRM and ticketing platforms
- Workflow agents coordinating actions across finance, HR, and operational systems
- Retrieval-Augmented Generation (RAG) pipelines accessing internal data repositories

- AI-driven analytics querying multiple internal services simultaneously

In many organizations, these agents function as **machine employees**, performing tasks that previously required human coordination between systems.

Unlike traditional users, however, AI agents often operate at machine speed and interact with numerous services simultaneously. A single automated workflow may involve multiple APIs, internal applications, and data systems interacting in sequence.

As organizations expand their use of AI-driven automation, these interactions dramatically increase the number of **internal system relationships** inside the enterprise.

In other words, AI does not simply introduce new identities.

It introduces **entirely new patterns of communication across enterprise infrastructure**.

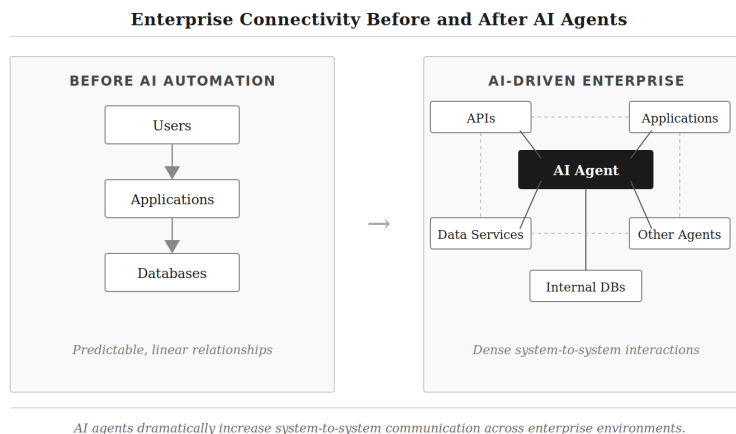


Figure 1: AI agents dramatically increase system-to-system communication across enterprise environments.

This shift represents a fundamental change in how enterprise systems interact.

2. AI Agents Expand the Internal Attack Surface

As AI agents increase automation across enterprise environments, they also expand the internal attack surface. Agents frequently require access to multiple systems in order to perform their tasks. They may retrieve information from internal databases, write updates to applications, trigger workflows across platforms, or coordinate actions across multiple services.

Each of these interactions creates a **system-to-system communication path** within the enterprise environment.

Many of these communication paths are:

- Newly created
- Dynamically evolving
- Not previously documented in security policy
- Difficult to review manually

Each connection creates a **trust relationship between systems**. As the number of these relationships grows, so does the organization's **internal trust surface**.

The internal attack surface is ultimately defined by the internal trust surface.

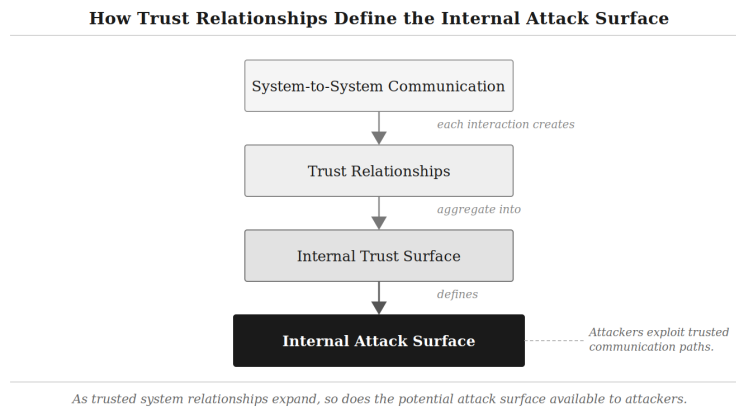


Figure 2: As trusted system relationships expand, so does the potential attack surface available to attackers.

Attackers rarely move laterally by bypassing trust relationships. They exploit them.

They leverage legitimate communication paths between systems to expand their access inside the network. When automation and AI increase the number of trusted connections between systems, they also increase the potential pathways that attackers can exploit after an initial compromise.

This dynamic does not necessarily involve malicious traffic entering the network. Instead, it emerges from legitimate systems communicating in ways that were never intentionally designed or reviewed.

In many cases, the greatest risk comes from **unintentional trust relationships** - system connections that exist simply because nothing was governing whether they should. As AI-driven automation expands these relationships, the **blast radius of compromise** within enterprise environments can grow significantly.

3. Why Existing Security Controls Fall Short

Enterprise security architectures already include multiple layers of protection. Identity platforms govern authentication, segmentation restricts network access, and detection systems monitor activity for signs of compromise. These controls remain essential. However, they were largely designed for environments where system relationships evolved slowly and could be defined in advance. AI-driven automation challenges these assumptions.

Identity Governance Does Not Govern System Relationships

Identity and Access Management systems are designed to answer a specific question: Is this identity allowed to access this resource? IAM governs authentication, authorization, and identity lifecycle management. These capabilities are critical for controlling access to enterprise systems.

However, IAM does not determine whether the **system-to-system relationships created by those identities are appropriate or necessary**. An AI agent may authenticate successfully to multiple systems and interact with several services during an automated workflow. Each individual access request may be fully compliant with IAM policy - yet the resulting system interactions may never have been intentionally designed, reviewed, or governed.

Traditional Segmentation Models Cannot Keep Pace

Network segmentation is commonly used to enforce predefined trust boundaries within enterprise environments. Security teams typically observe traffic patterns, define segmentation policies, and deploy enforcement rules that restrict communications between systems. This model works best when communication patterns are relatively stable.

AI-driven automation introduces continuously evolving workflows and integrations. Agents may trigger new API interactions, access additional services, or expand automation across different systems over time. As automation accelerates, segmentation policies increasingly reflect yesterday's infrastructure rather than today's reality.

Organizations often face two difficult options:

- Allow broader communication zones to avoid disrupting automation
- Continuously rewrite segmentation policies to keep pace with changes

Both approaches introduce operational challenges and can weaken the effectiveness of segmentation controls over time.

Detection Systems Face Increasing Signal Noise

Detection platforms such as Network Detection and Response (NDR) and Extended Detection and Response (XDR) monitor environments for suspicious activity and anomalous behavior. AI-driven automation can dramatically increase the volume of legitimate internal activity within enterprise networks.

Automated workflows may generate:

- New communication paths between systems
- Increased east-west network traffic
- Automated API interactions
- Unfamiliar patterns of service-to-service communication

From the perspective of detection systems, many of these behaviors may appear anomalous. Security teams are therefore required to determine whether alerts represent malicious activity or legitimate automation. As AI adoption increases, the volume of anomalous but legitimate activity can grow significantly, increasing the potential for investigation overhead and alert fatigue within security operations teams.

Visibility Into AI Traffic Is Necessary - But Not Sufficient

Many organizations are now asking whether they can identify AI-driven activity inside their networks - including RAG pipelines, agent orchestration traffic, and other AI-related communications. This visibility is important. Security teams must understand where AI systems are operating and what systems they interact with.

However, identifying AI traffic alone does not address the broader challenge created by AI-driven automation. AI systems frequently interact with multiple internal services, databases, and applications as part of their workflows. These interactions can introduce new system-to-system communication paths that were never explicitly designed or reviewed.

**The goal should not simply be detecting AI traffic.
It should be governing the trust relationships AI systems create inside the enterprise.**

4. The Missing Layer: Internal Trust Governance

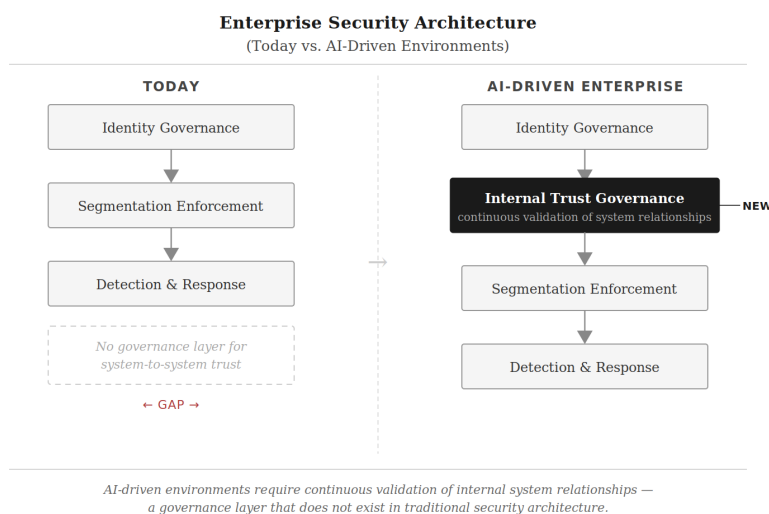


Figure 3: AI-driven environments require continuous validation of internal system relationships - a governance layer that does not exist in traditional security architecture.

These dynamics reveal an emerging gap in enterprise security architecture. Existing security controls focus on authentication, policy enforcement, and anomaly detection. But they do not continuously evaluate whether internal system relationships themselves are appropriate and safe.

As AI-driven automation expands, organizations increasingly need the ability to answer a new set of questions:

- Should these systems be communicating at all?
- Is this communication necessary for the systems involved to perform their intended roles?
- Is the behavior consistent with the purpose of the assets involved?

Answering these questions requires a capability that goes beyond traditional segmentation or detection. It requires **continuous governance of internal communications** across enterprise infrastructure.

In practice, continuous internal trust governance means maintaining a dynamic, validated model of which system-to-system communications should exist across the enterprise - and continuously evaluating observed communications against that model. Rather than defining static segmentation policies in advance, this approach derives expected communication behavior from the purpose and role of each asset in the environment. When a new communication path appears - whether introduced by an AI agent, an integration, or an attacker exploiting a legitimate pathway - it can be evaluated immediately against what is known to be appropriate for those systems. Relationships that are unnecessary, anomalous, or inconsistent with the intended role of the assets involved can be flagged or interrupted before they

become a liability. The result is a living trust model that evolves alongside the environment rather than falling behind it.

5. The Next Evolution of Zero Trust

Over the past decade, Zero Trust principles have reshaped enterprise security by eliminating implicit trust at the perimeter. Identity systems now continuously verify users and enforce strict access controls before granting access to resources.

However, modern digital infrastructure extends far beyond user-to-application interactions. Today's enterprise environments include cloud services, APIs, automation platforms, AI systems, and countless internal services communicating continuously with one another. As these environments become increasingly interconnected, security architectures must evolve to address not only **who is accessing systems**, but also **how systems themselves interact**.

This shift represents the next stage in the evolution of Zero Trust.

In addition to governing identity-based access, organizations must establish discipline around **internal system communications** - ensuring that system relationships are intentional, necessary, and safe.

IAM governs trust for users and agents.

The next challenge is governing trust between the systems they connect.

New security approaches are beginning to address this challenge by introducing continuous validation of internal communications and system relationships.

This emerging discipline - governing system-to-system trust across the enterprise - is Zero Trust for Internal Networks, ZTNX, extending Zero Trust principles beyond identity verification to the governance of internal system trust.

The proliferation of AI agents will dramatically increase the number of machine identities operating inside enterprise environments. As automation expands, the trust surface of machine-to-machine communication will grow far faster than the human access surface.

When that happens, internal trust governance becomes inevitable.
