xage SECURITY

WHITEPAPER

Operationalizing CISA's AI-in-OT Principles: Zero Trust Enforcement with Xage

Introduction

AI is moving from "pilot" to "plant" across critical infrastructure. In OT, that shift is not a normal IT modernization story, it is a safety, reliability, and national resilience issue. The same models that can improve predictive maintenance and accelerate troubleshooting can also introduce novel failure modes (drift, unsafe recommendations, opaque decision chains) and new cyber pathways into environments built for determinism and containment.

That's why the CISA's new joint guidance—**"Principles for the Secure Integration of Artificial Intelligence in Operational Technology,"** published December 3, 2025—matters. It is not hype. It is an operational blueprint from CISA and peer agencies across the U.S. and allied nations, written explicitly for critical infrastructure owners and operators.

The guidance frames secure AI-in-OT in four principles: **1) Understand AI, 2) Consider AI Use in the OT Domain, 3) Establish AI Governance and Assurance Frameworks, and 4) Embed Safety and Security Practices into AI and AI-enabled OT systems.**

OT security leaders face the immediate challenge of translating CISA's AI principles into enforceable controls. These controls must be effective despite legacy system constraints, the reality of vendor access, and intermittent connectivity. Zero Trust offers a practical architectural solution to this challenge, moving beyond buzz words. It ensures continuous verification, auditability, and rapid containment, making certain that both AI components and their human operators are strictly limited to their explicitly authorized actions.

The Xage platform is specifically designed for critical infrastructure environments. It delivers identity-first access control and policy enforcement across OT/IT systems, offering solutions like secure remote access, privileged access management, segmentation, and tamper-resistant auditability. Critically, it is engineered to function reliably without requiring continuous cloud connectivity.

Below is a practitioner-focused mapping of each principle to concrete implementation patterns, and to the Xage capabilities that operationalize them.

Principle 1: Understand AI

The joint guidance is direct: treat AI as a distinct risk domain inside OT, with unique failure and abuse cases—including drift over time and safety-process bypasses—and manage those risks as part of availability and reliability engineering. It calls out three practical expectations:

- Understand the unique risks and operational impacts of AI in OT environments.
- Apply a secure AI system development lifecycle (design → procurement → deployment → operations).
- Educate OT personnel so humans can interpret AI outputs, recognize failure modes, and keep manual competencies intact.

How Xage Assists With Principle 1

Principle 1 is where many teams underestimate the "integration tax." In OT, "understanding AI" is not just model literacy, it is system boundary clarity: what the AI can reach, what it can change, and how you prove it.

Xage helps by turning AI from an opaque component into a governed identity with constrained privileges:

- **Treat AI agents as first-class identities (not anonymous processes).** Xage's architecture is identity-centric—extending Zero Trust controls to users, devices, services, and non-person entities—so AI workloads can be onboarded with explicit identities and scoped permissions rather than broad network trust.
- **Define and enforce least-privilege interaction paths.** Instead of "AI can talk to the OT network," policies can be written so the AI can only read specific telemetry sources, or can only reach a brokered interface, or can only act through controlled workflows. This is foundational to understanding, because it makes "what the AI is allowed to do" deterministic and reviewable.
- **Build understanding through evidence (not assumptions).** Xage emphasizes auditability and traceability for access and activity across protected resources, which supports the operational need to validate what AI components are actually doing once deployed.

In OT, "understanding AI" becomes practical when AI is constrained by identity and policy, not by trust in model behavior.

Principle 2: Consider AI Use in the OT Domain

Principle 2 emphasizes disciplined tradecraft before adoption, focusing on the business case, data security, and vendor realities of AI use in OT environments.

- **Assess the OT business case** for AI and whether simpler alternatives can meet the need.
- **Manage OT data security risks,** including where data goes, who accesses it, and how compromise impact is minimized.
- Be explicit about the **role of OT vendors** embedding AI, including connectivity and data handling expectations.

A specific recommendation is especially important for architecture: prefer push-based or brokered patterns that move required features or summaries out of OT without granting persistent inbound access, so the AI system does not become a standing attack path into OT.

How Xage Assists With Principle 2

Principle 2 is where OT programs succeed or fail based on how they handle data flow and third-party access. Xage enables several practical patterns aligned to the guidance:

- **Push/brokered architectures instead of inbound reach.** Xage is designed as an overlay security fabric that supports brokered access and segmentation.
- **Secure data movement as a controlled exchange.** Xage's Zero Trust Data Exchange is explicitly positioned around protecting data flows, addressing integrity and controlled exchange needs that show up repeatedly in AI data pipelines.
- **Vendor access without expanding the blast radius.** The guidance highlights vendor considerations because vendors increasingly embed AI features that may require new connectivity. Xage's secure remote access and privileged access controls provide a way to grant tightly scoped, time-bounded access to specific resources, without turning on broad VPN-style connectivity.
- **Operate securely even when connectivity is imperfect.** The guidance is written for real critical infrastructure. Xage emphasizes architectures that continue operating without assuming cloud dependency, including deployments designed for degraded or disrupted connectivity.

Principle 2 is fundamentally about minimizing new attack paths and controlling data exposure. Xage supports that by enabling brokered access, controlled data exchange patterns, and constrained vendor interactions.

Principle 3: Establish AI Governance and Assurance Frameworks

The joint guidance positions governance and assurance as ongoing processes rather than one-time reviews, emphasizing AI systems that are auditable, testable, and continuously evaluated.

- **Implement governance mechanisms** (roles, responsibilities, accountability) for AI in OT.
- **Integrate AI** into existing security frameworks and processes (audit, vuln management, incident response).
- Conduct thorough **testing and evaluation** before production use, then continuously reassess.

How Xage Assists With Principle 3

Governance fails when policies live in documents and exceptions live in the network. The practical value of Xage here is that it provides a **policy enforcement plane** and a **consistent control surface** across environments:

- **Policy-as-enforcement, not policy-as-intent.** Xage's model centralizes policy definition and pushes enforcement into the environment so governance decisions (who can access what, under what conditions) become technical reality, not a "best effort."
- **Assurance through repeatable controls across test and production.** The guidance stresses testing and evaluation; the easiest way to make testing meaningful is to keep access controls consistent between environments. A platform-based policy model supports that by allowing governed patterns to be validated in a representative environment before rollout.
- **Privileged activity controls that stand up to audit.** OT AI programs will raise auditor questions fast: "Who approved this access?" "What did the system change?" "Can you prove the integrity of the record?" Xage's privileged access is designed to bring structure and traceability to high-risk operations.

Principle 3 is where organizations separate "AI experimentation" from "AI operations." Xage supports that shift by making AI access governable, testable, and audit-ready.

Principle 4: Embed Oversight and Failsafe Practices Into AI and AI-Enabled OT Systems

This principle is the operational heartbeat of the guidance, embedding oversight and fail-safes to ensure humans remain accountable and systems fail safely.

- Maintain **inventory and visibility** into AI components.
- Establish monitoring and oversight mechanisms, and **log and monitor AI inputs/outputs.**
- **Limit active control of OT infrastructure by AI without a human in the loop,** and ensure AI actions are distinguishable via identity in logs.
- Prefer architectures that **avoid persistent inbound access** from AI systems into OT, so AI is not a standing attack path.
- Embed **failsafe mechanisms** and incorporate AI failure states into incident response and functional safety procedures.

How Xage Assists With Principle 4

Principle 4 is where OT leaders want more than recommendations, they need mechanisms that can interrupt unsafe paths quickly and prove what happened after the fact. Xage enables several enforceable safeguards:

- **Human-in-the-loop control via policy and privileged workflows.** If AI outputs can influence operations, Xage can constrain those actions behind privileged access controls and approvals, supporting a design where AI can recommend, but humans authorize execution for safety-critical operations.
- **Containment by default through segmentation and least privilege.** If an AI component (or its supporting infrastructure) is compromised, the goal is containment, not cleanup after lateral movement. Xage's segmentation and identity enforcement model is designed to limit blast radius by preventing implicit trust and restricting interaction paths.
- **Operational visibility and forensic support.** The guidance stresses monitoring, logging, and distinguishing AI identities in audit trails. Xage's focus on identity-based access decisions and traceability supports the operational need to show which identity did what, and to integrate that evidence into incident response.
- **Failsafe readiness as an architectural feature.** Because access is policy-mediated, it is feasible to define "safe mode" behaviors (e.g. disabling AI write access globally, restricting to read-only telemetry, or revoking a specific AI identity) without redesigning the OT network during an incident. This aligns directly to the guidance's requirement to incorporate AI failure states into response and safety processes.

Principle 4 is not optional in OT. Xage supports it by making AI influence paths constrainable, monitorable, and reversible under defined safety and IR procedures.

xage SECURITY

What to Do Next: A Practical, 30-60 Day Plan OT Leaders Can Execute

For most operators, the fastest responsible path is not "deploy AI everywhere." It is to pick one bounded use case and implement the guardrails as a reusable pattern:

- **Define the AI system boundary:** components, data sources, where inference runs, where outputs go. (Make it inventory-grade.)
- **Adopt push/brokered connectivity:** move required data/features out of OT; avoid persistent inbound access paths.
- **Assign AI identities and least privilege:** treat AI agents/services like privileged entities with explicit permissions and auditable paths.
- **Implement human-in-the-loop gates for safety-relevant actions:** AI can recommend; humans approve changes that can affect process safety.
- **Bake in failsafe states and IR steps:** define how to bypass/replace AI, and how to respond to AI-targeted malicious activity or AI failure.

This is where a Zero Trust platform earns its keep: transforming these steps into controls that are applied consistently across sites, vendors, and changing operational conditions.

Cisa's Principles Are Actionable, If You Make Them Enforceable

The joint guidance is clear-eyed: AI can drive efficiency and better decision-making for critical infrastructure, but it also introduces new risks that must be managed to protect the safety, security, and reliability of OT environments.

For OT security leaders, the way forward is not to slow innovation, it is to govern it. That means treating AI as a privileged participant in OT: explicitly identified, minimally permitted, continuously monitored, and designed with human oversight and failsafe states from day one.

Xage aligns naturally to that outcome because it is built as a critical-infrastructure Zero Trust platform: **policy-enforced access, segmentation and containment, privileged activity controls, and operational resilience**—capabilities that map directly to what the guidance is asking operators to implement.

If we treat these principles as engineering requirements, not as a compliance checkbox, we can adopt AI in OT or any other environment in a way that scales safely: bounded by policy, proven by evidence, and resilient under real-world conditions.