Healthcare AI Governance and Security with Cognome on Lenovo ThinkSystem

Solution Brief

Executive Summary

Healthcare organizations expanding AI across clinical, operational, and administrative workflows need to discover AI in use, govern approved models, and understand how those models produce their outputs. A model registry establishes an inventory, monitoring reveals behavior over time, and explainability provides transparency and traceability. Effective oversight requires these capabilities to work together.

Cognome connects AI Sniffer™ for agentless AI discovery and security, ExplainerAI™ for model intake, continuous monitoring, and clinician-level explainability, and SOLUS™ for correlated healthcare AI risk intelligence. Together, they help health systems identify sanctioned and shadow AI, protect sensitive information, monitor Responsible AI Use including drift, bias, and hallucinations, and maintain evidence for governance reviews and audits.

Deployed on Lenovo ThinkSystem infrastructure powered by Intel® Xeon® 6 processors, the solution runs on premises or in the customer's private cloud with no Cognome SaaS component. CPU-based inference supports ExplainerAI workloads, with GPU configurations available when warranted. Lenovo compute, memory, storage, and support services provide a scalable foundation for AI oversight while keeping governance data within the customer's environment.

Business Challenge

AI adoption is outpacing traditional healthcare governance practices. Clinical leaders, cybersecurity teams, and IT administrators need to know which AI tools are active, how production models behave, and whether their outputs can be understood and reviewed.

- **Unseen AI usage:** Local models, embedded vendor AI, and unauthorized tools can operate outside central inventories. Biomedical and medical IoT devices may not support endpoint agents, creating additional discovery gaps.
- **Incomplete model oversight:** A model registry records what has been approved but does not provide continuous monitoring or explainability. Health systems need a connected process for model intake, performance oversight, and transparent review of outputs.
- **Sensitive data exposure:** Everyday actions, such as pasting patient information into a public chatbot, can expose protected health information and create privacy risks.
- **Evolving model behavior:** Performance degradation, drift, bias, and hallucinations can emerge after deployment, requiring ongoing monitoring and investigation.
- **Fragmented evidence and ownership:** Records spread across teams make governance reviews difficult, while abandoned AI without a clear owner can remain active and unmanaged.
- **Gaps in existing security coverage:** Endpoint, network, and Microsoft security tools serve important roles, but may not identify AI-specific activity or provide clinical model governance.

Health systems need coordinated AI discovery, monitoring, and explainability that works alongside existing security tools and keeps sensitive governance data within their environment.

Solution Overview

Lenovo and Cognome combine enterprise infrastructure with an integrated healthcare AI governance and security platform. AI Sniffer™ discovers AI across the network without requiring endpoint software. Discovered models can move into ExplainerAI™ through a guided onboarding wizard, pre-filled from their discovery records, connecting discovery with model intake and continuous monitoring. SOLUS™ correlates discovery and governance findings into healthcare AI risk intelligence. The platform runs on premises or in the customer's private cloud with no Cognome SaaS component.

Table 1. Solution overview

Component	Primary Role	Customer Value
AI Sniffer™	Discover and inventory AI through agentless network discovery, with optional agents for deeper endpoint visibility and PHI protection.	Visibility into sanctioned and shadow AI, including devices that cannot carry agents.
ExplainerAI™	Connect model intake and registry with continuous monitoring and clinician-level explainability.	Ongoing clinical AI oversight, alerts, and governance evidence.
SOLUS™	Correlate discovery and model governance findings in a healthcare AI risk intelligence layer.	Unified risk review, framework mapping, and audit support.
Lenovo SR630 V4 and SR650 V4 Servers	Host Cognome's discovery, analytics, and model governance workloads.	Scalable enterprise infrastructure within the customer's environment.

AI Discovery, Security and Model Governance

Cognome connects AI discovery, model intake, monitoring, and explainability in one governance workflow. Health systems can identify previously unknown AI, onboard models for oversight, and correlate findings across clinical, security, and compliance teams.

Discover AI across the environment

AI Sniffer discovers AI across networks without requiring software installation on the devices being inspected. Network scans, file inspection, and discovery sources such as DNS logs help identify local models, vendor-embedded AI, and shadow AI, including biomedical and medical IoT devices that cannot carry an agent. Cognome works alongside existing endpoint, network, and Microsoft security tools to provide AI-specific visibility.

Optional AI Sniffer Agents add deeper endpoint visibility and PHI protection. Agents classify clipboard content as users interact with AI tools, helping identify patient information being pasted into public chatbots. Cognome detects 16 of the 18 HIPAA identifiers, with sensitive-content classification performed locally. Endpoint controls support quarantine or blocking of endpoints running unauthorized AI.

Connect discovery to model governance

Models discovered by AI Sniffer can be onboarded into ExplainerAI through a guided wizard pre-filled from their discovery records. This connects discovery with model registry and intake, helping teams establish ownership and configure ongoing monitoring.

ExplainerAI monitors onboarded models for Responsible AI Use, including drift, bias, hallucinations, and performance changes. Model-specific thresholds and centralized alerts support investigation, while clinician-level explainability provides transparency and traceability for reviewing model outputs.

Correlate healthcare AI risk

SOLUS brings together discovery findings and model governance evidence to support coordinated risk review. It maps findings to HIPAA, NIST AI RMF, HSCC AI Governance, and NIH Responsible AI, helping teams maintain evidence for governance reviews and audit preparation.

Discovery also helps identify abandoned AI that lacks an accountable owner, supporting investigation and retirement of unmanaged assets.

Integration with Healthcare Systems

ExplainerAI integrates with Epic and other EHR environments, with Cognome's AutoETL connector supporting deeper Epic monitoring and analytics. Customer planning establishes the required interfaces, data access, HL7 message volumes, and monitoring feeds.

Coverage of AI agents and MCP servers is in development.

Table 2. Core capabilities and operational value.

Capability	Operational value
Agentless AI discovery	Identify sanctioned and shadow AI, including devices that cannot carry agents.
Endpoint PHI protection	Classify sensitive content locally and identify risks during user interactions with AI tools.
Guided model intake	Identify drift, bias, hallucinations, output anomalies, and changes in model behavior for further review.
Clinical explainability	Carry discovery information into model registration and monitoring setup.
Responsible AI Use monitoring	Identify drift, bias, hallucinations, and performance changes for investigation.

Capability	Operational value
Clinician-level explainability	Provide transparency and traceability to support review of model outputs.
Endpoint controls and alerts	Support quarantine, blocking, and investigation of unauthorized AI activity.
SOLUS risk intelligence	Correlate findings and map evidence to healthcare requirements and AI risk frameworks.

Target Healthcare Use Cases

Healthcare organizations can apply Cognome's integrated platform to discover AI, protect patient information, and govern models throughout their operational lifecycle.

The following table lists some use cases:

Table 3. Target healthcare use cases.

Healthcare example	What the solution enables	Illustrative success measure
Shadow AI across hospitals and clinics	Discover unauthorized tools, local models, and vendor-embedded AI without requiring endpoint agents, including biomedical and medical IoT devices.	AI assets discovered; previously unknown assets reviewed.
Potential PHI exposure	Use AI Sniffer Agents to classify clipboard content locally and identify PHI exposure during interactions with public chatbots.	PHI-related events investigated and resolved; time to investigation.
Onboarding discovered models	Use discovery records to pre-fill guided model intake and establish monitoring in ExplainerAI.	Models onboarded; time from discovery to monitoring.
Clinical model performance and explainability	Monitor models such as sepsis predictors for drift, bias, and performance changes, with clinician-level explainability supporting transparent review of outputs.	Models monitored; alerts reviewed; explainability evidence available for review.
Abandoned or unowned AI	Identify unmanaged AI assets so teams can establish ownership, assess risk, and retire assets that are no longer needed.	Unowned assets reviewed, assigned, or retired.
Audit and governance oversight	Correlate discovery findings and model governance evidence through SOLUS to support risk assessments and audit preparation.	Evidence completeness; governance review and audit preparation effort.

Business Outcome

In the following sections, we take a deeper dive:

- Business Outcome
- Recommended Infrastructure Sizing
- Proposed Lenovo Platform Mapping

Business Outcome

Lenovo and Cognome help healthcare organizations expand AI adoption with clearer accountability, stronger patient-data protection, and more predictable operating costs.

- **More complete AI visibility:** Discover sanctioned, shadow, and abandoned AI across the enterprise, including devices that cannot carry endpoint agents.
- **Connected governance throughout the AI lifecycle:** Move discovered models through intake, registration, and ongoing monitoring, helping teams establish ownership and maintain oversight.
- **Stronger clinical AI accountability:** Combine Responsible AI Use monitoring with clinician-level explainability to support transparent, traceable review of model behavior and outputs.
- **Reduced exposure and incident impact:** Identify endpoint PHI risks and unmanaged AI, helping teams investigate incidents and retire abandoned assets that create ongoing exposure.
- **More efficient governance reviews:** Correlate inventory, alerts, and supporting evidence to streamline investigations, risk assessments, and audit preparation.

- **Data sovereignty:** Run on premises or in the customer's private cloud with no Cognome SaaS component, keeping governance data within the customer's environment.
- **More predictable costs:** Cognome's pricing has no consumption fees, while CPU-based ExplainerAI configurations provide an infrastructure option without dedicated GPUs for suitable workloads.

Recommended Infrastructure Sizing

AI Sniffer™ and ExplainerAI™ running on Lenovo infrastructure help healthcare organizations discover AI activity, govern deployed models, and turn findings into actionable risk information across network, endpoint, and edge environments.

Lenovo ThinkSystem SR630 V4 and SR650 V4 servers provide a scalable foundation for these workloads. The following table presents infrastructure configurations based on Cognome's AI Sniffer Server and ExplainerAI compute requirements.

Proposed Lenovo Platform Mapping

ThinkSystem SR630 V4 and SR650 V4 are proposed as the server platform family for Cognome's containerized workloads. Final platform selection, physical host count, CPU/RAM population, GPU compatibility, network design, and high-availability layout require a Lenovo–Cognome design review.

Table 4. Lenovo Infinidat Systems with Veeam Ready validation.

Sizing	Endpoints Supported	Governed Models	Physical Servers	CPU Config	Total CPU Cores	RAM Config	Total Memory
Small	Up to 1,000	Up to 5	1	2 x 32 core	64	16 x 32 GB	512 GB
Medium	Up to 5,000	Up to 25	1	2 x 64 core	128	16 x 64 GB	1024 GB
Large	Up to 25,000	50+	2	2 x 64 core per server	256	16 x 64 GB per server	2048 GB

Note: Configurations use Lenovo ThinkSystem SR630 V4 or SR650 V4 servers. Memory follows the recommended balanced DIMM population of eight DIMMs per CPU, providing capacity above Cognome's stated requirements as headroom. CPU core counts and total memory represent the combined capacity across all physical servers listed.

Endpoint count determines AI Sniffer sizing, while governed model count determines ExplainerAI sizing. Each capability should be sized according to customer requirements. Extra-large deployments require custom sizing. Final configurations depend on workload, deployment, and availability requirements.

Solution Architecture

The solution connects agentless AI discovery, optional endpoint protection, and production-model governance on Lenovo ThinkSystem infrastructure. AI Sniffer™ discovery records support guided onboarding into ExplainerAI™, linking model intake and registration with continuous monitoring and clinician-level explainability. SOLUS™ correlates discovery findings and governance evidence into unified healthcare AI risk intelligence.

The diagram below provides a conceptual view of the design and solution flow.

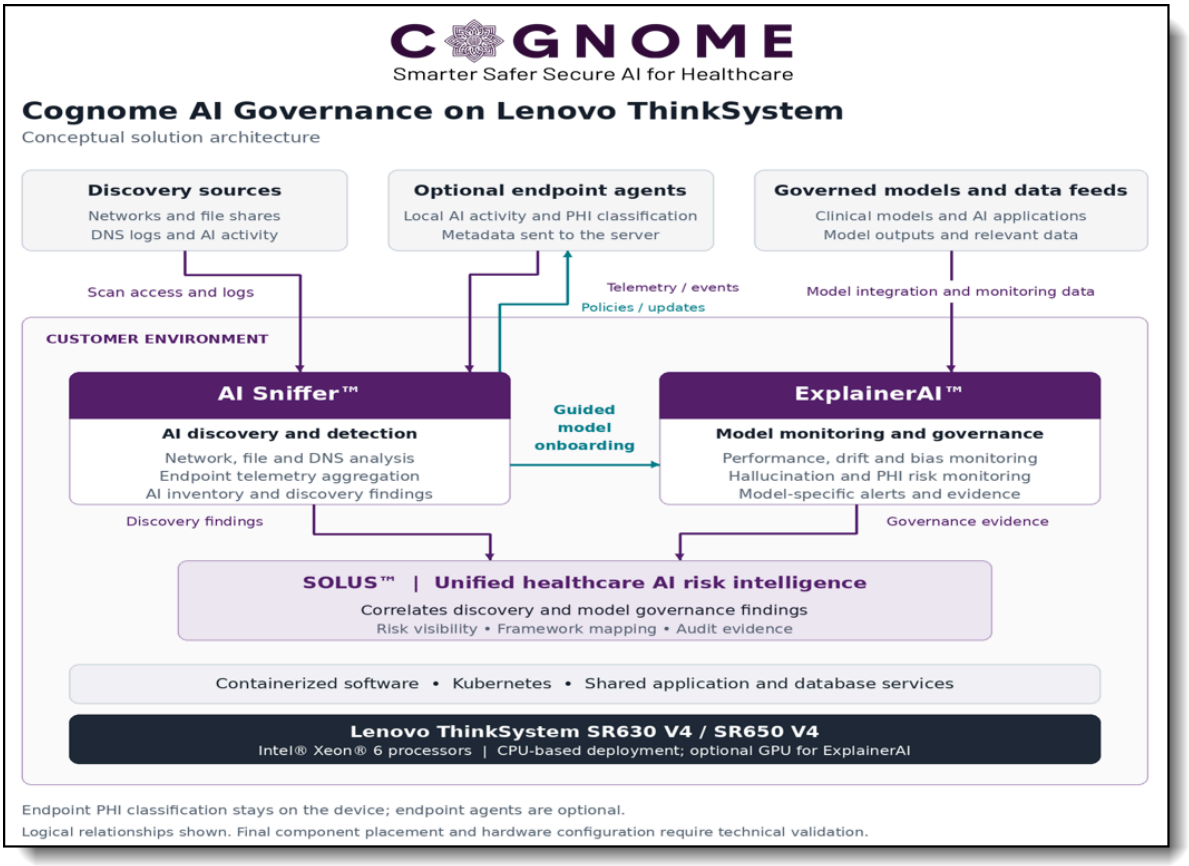


Figure 1. Conceptual Cognome AI discovery and governance on Lenovo ThinkSystem

Infrastructure and Integration Considerations

Cognome's server-side services run as containers orchestrated by Kubernetes within the customer's environment. Combined deployments share application and database services, while AI Sniffer ingest capacity and ExplainerAI inference resources are sized independently. Optional endpoint agents classify sensitive content locally, send metadata and events to the discovery server, and receive policies and updates through secure, bidirectional communication.

The platform runs on premises or in the customer's private cloud with no Cognome SaaS component, keeping governance data within the customer's environment. Healthcare integrations provide the model and clinical data needed for monitoring, while SOLUS maps findings to HIPAA, NIST AI RMF, HSCC AI Governance, and NIH Responsible AI.

Table below lists the Infrastructure and integration considerations

Table 5. Infrastructure and integration considerations

Planning area	Deployment information to confirm
AI discovery	Subnets / CIDR ranges, network scan permissions, available discovery sources, including DNS logs, accessible file shares, number of endpoints and enrolled agents.
Model governance	Models approved for monitoring, integration points, HL7/feed rate, data rights, inference volume, refresh frequency, alert thresholds.
Software platform	Linux x86-64, Kubernetes 1.27+, container runtime, identity integration, agent enrollment and update approach.
Network / storage	Cognome recommends 10 GbE between nodes (1 GbE may be acceptable at the smallest size); usable storage excludes resilience overhead.
Operations and availability	Customer backup, storage redundancy, infrastructure monitoring, disaster recovery; add nodes when control-plane HA is required.

Cognome notes: Configurations below three Kubernetes control-plane nodes do not provide control-plane high availability, and a node failure may require manual recovery. Requirements for enhanced availability and security hardening should be established before a production design is approved.

Getting Started

Organizations can begin with Cognome's Shadow AI Discovery or Shadow AI Risk Assessment engagement to establish AI visibility and identify governance priorities. An initial AI inventory can be available in days, depending on scope. Lenovo and Cognome then work with the customer to define model onboarding, healthcare integrations, security requirements, and infrastructure sizing for ongoing governance.

Related Links

For more information, see these resources:

- For more information about Lenovo solutions, contact your Lenovo representative or Business Partner, or visit Lenovo servers.
- For more information about Cognome's AI discovery, security, and governance capabilities, visit Cognome

Product features, validation status, and availability should be checked with Lenovo and Cognome before customer publication.

Authors

Erik Robert is an AI Ecosystem and Solutions professional focused on identifying, developing, and bringing enterprise AI solutions to market through collaboration with independent software vendors (ISVs), infrastructure teams, and strategic technology partners. His work centers on evaluating emerging AI technologies and aligning them with scalable reference architectures across compute, storage, and software platforms. With experience spanning solution strategy, platform positioning, and AI go-to-market initiatives, Erik's goal is to help organizations adopt practical, production-ready AI solutions that deliver measurable business value.

Abed Islam is a Solutions Architect focused on designing scalable, practical AI platforms for enterprise and edge environments. His work centers on integrating compute, storage, networking, and AI software stacks through close collaboration with technology partners and industry providers. With technical experience in infrastructure architecture and deployment, Abed's goal is to make advanced AI capabilities accessible, reliable, and production-ready for real-world use.

Related product families

Product families related to this document are the following:

- AI Servers
- ThinkSystem SR630 V4 Server
- ThinkSystem SR650 V4 Server

Notices

Lenovo may not offer the products, services, or features discussed in this document in all countries. Consult your local Lenovo representative for information on the products and services currently available in your area. Any reference to a Lenovo product, program, or service is not intended to state or imply that only that Lenovo product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any Lenovo intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any other product, program, or service. Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

Lenovo (United States), Inc.
8001 Development Drive
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO PROVIDES THIS PUBLICATION ”AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary. Lenovo may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any references in this publication to non-Lenovo Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk. Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

© Copyright Lenovo 2026. All rights reserved.

This document, LP2584, was created or updated on November 20, 2026.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
  https://lenovopress.lenovo.com/LP2584
- Send your comments in an e-mail to:
  comments@lenovopress.com

This document is available online at https://lenovopress.lenovo.com/LP2584.

Trademarks

Lenovo and the Lenovo logo are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. A current list of Lenovo trademarks is available on the Web at https://www.lenovo.com/us/en/legal/copytrade/.

The following terms are trademarks of Lenovo in the United States, other countries, or both:
Lenovo®
ThinkSystem®

The following terms are trademarks of other companies:

Intel®, the Intel logo and Xeon® are trademarks of Intel Corporation or its subsidiaries.

Linux® is the trademark of Linus Torvalds in the U.S. and other countries.

Microsoft® is a trademark of Microsoft Corporation in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.