Red Hat

Layered security in the age of AI and emerging threats

By Lucy Huh Kerner,
Director, Security Global Strategy and Evangelism, Red Hat

Table of contents

The new security frontier **3**

The vulnerability gap: Why traditional IT operations fail against AI 3

Security is continuous at every layer 4

Security pillar 1: Build a strong security foundation in the age of AI **5**

Recommendations and best practices 5

Tactical steps 7

Security pillar 2: Protect the pipeline (Securing the software and AI supply chains) **8**

Recommendations and best practices 8

Tactical steps 9

Security pillar 3: Implement security, safety, and governance for agentic AI **11**

Recommendations and best practices 11

Tactical steps 12

Security pillar 4: Automating security at scale for machine-speed defense **13**

Recommendations and best practices 13

Tactical steps 15

Ready to get started? **16**

In 2025, there was a

66.6%

increase in CVEs compared to 2023.[1]

45%

of discovered vulnerabilities in large organizations remain unpatched after 12 months.[2]

79%

of organizations agree that Generative AI creates entirely new security challenges.[3]

63%

of organizations currently lack AI governance policies.[4]

The new security frontier

Modern organizations are navigating an increasingly complex security landscape—balancing risk mitigation with regulatory compliance, digital sovereignty, supply chain security, and AI security across hybrid cloud environments. At the same time, the rise of AI-driven workloads, zero trust architectures, automation, and the need to prepare for post-quantum cryptography are reshaping how security is approached at every level.

The modern enterprise is navigating a profound shift in digital defense. Historically, the "defensive perimeter" served as the cornerstone of security—a model where firewalls acted as binary gatekeepers. Under this framework, anything inside the network was implicitly trusted, while outside actors were blocked.

However, the rise of AI and the impending arrival of quantum-ready threats have rendered this strategy obsolete. As AI-powered threats fundamentally alter enterprise vulnerabilities, organizations must transition from a reactive security posture to a model of proactive, automated resilience. This evolution requires a fundamental shift: moving away from viewing security as a point-in-time event and instead treating it as a continuous discipline.

This discipline is structured around a lifecycle of identifying assets, protecting operations, detecting anomalies quickly, responding with coordinated precision, and recovering services to improve long-term resilience. Crucially, it involves the mitigation of future risks, such as the "harvest now, decrypt later" strategy employed by actors targeting data today to exploit with tomorrow's quantum computing capabilities.

Modern enterprise security requires a layered security strategy that makes sure that a failure in 1 layer does not lead to a systemic breach. This e-book provides a blueprint for implementing a layered security defense, from the operating system (OS) foundation to autonomous AI agents, maintaining enterprise resilience in the age of AI and emerging threats.

The vulnerability gap: Why traditional IT operations fail against AI

In the current era of hyper-accelerated IT operations, mean time to remediation (MTTR) has emerged as the primary metric for organizational response. In an environment where AI-driven tools can scan for weaknesses, generate exploits, and launch attacks across a global surface in seconds, traditional, manual patching cycles are no longer just inefficient, they are a liability. When remediation lags, the risk to enterprise stability, developer productivity, and brand reputation grows. For the modern Chief Information Security Officer (CISO), closing this gap should be a core area of focus to make sure that digital transformation does not result in digital fragility. A proactive stance on security and remediation is a strong defense against the scale of modern threats.

1 "Published CVE Records." CVE Program, accessed 31 July 2026.

2 Keary, Eoin. "The Vulnerability Backlog Crisis: Why 45% of Enterprise Vulnerabilities Never Get Fixed," Edgescan blog, 21 Oct. 2025.

3 Red Hat e-book, "The state of cloud native security report: 2026 edition." 22 Dec. 2026.

4 "Cost of a Data Breach Report 2025." IBM, accessed 31 July 2026.

The rapid democratization of AI has handed attackers a force multiplier. Today, AI models can scan millions of lines of code to find zero-day vulnerabilities or automate the weaponization of newly disclosed Common Vulnerabilities and Exposures (CVEs) in seconds. Traditional IT operations, which rely on manual triage and human-centric workflows, are structurally incapable of keeping pace with this machine-speed threat landscape. This creates a vulnerability gap, meaning a window of exposure that grows wider as enterprise complexity increases.

In 2025, more than 48,000 CVEs were reported, making a 66.6% increase in CVEs compared to 2023.[1] Expecting internal teams to manually triage and maintain this volume while also delivering critical AI initiatives is not a viable strategy. In fact, 45% of discovered vulnerabilities in large organizations remain unpatched after 12 months.[2]

In addition, 79% of organizations agree that generative AI creates entirely new security challenges.[3] Among the largest of those challenges is establishing strong governance. Many organizations do not know if sensitive data is being exposed to insecure AI tools, and 63% of organizations currently lack AI governance policies.[4] The rapid adoption of agentic AI tools is far outpacing the creation of formal policies and documentation.

Security is continuous at every layer

In an era where exploits are weaponized by the very technology many organizations aim to deploy, an AI-ready enterprise is only as resilient as its weakest architectural layer. A disconnected approach to security cannot scale at the pace of technology innovation, putting critical data at constant risk.

An AI-ready enterprise requires a layered defense. The entire stack needs to be an integrated solution, where security and resilience are integrated into the fabric of the infrastructure, applications, and AI. An enterprise AI journey should be built on a foundation that is secured by design and automated by default.

To build a layered, defensive, and resilient security architecture:

- **Build on a strong security foundation.** Safeguard the underlying Linux® OS and container platforms to maintain a trusted environment for AI workloads.
- **Protect the pipeline.** Integrate automated security standards throughout the software and AI supply chains and lifecycle, from upstream open source to production.
- **Secure the intelligence.** Implement security, safety, and governance for agentic AI.
- **Automate security at scale by default.** Transform security from a manual compliance burden into automated means for scaling AI implementation.

1 "Published CVE Records." CVE Program, accessed 31 July 2026.

2 Keary, Eoin. "The Vulnerability Backlog Crisis: Why 45% of Enterprise Vulnerabilities Never Get Fixed," Edgescan blog, 21 Oct. 2025.

3 Red Hat e-book, "The state of cloud native security report: 2026 edition." 22 Dec. 2026.

4 "Cost of a Data Breach Report 2025." IBM, accessed 31 July 2026.

Security pillar 1: Building a strong security foundation in the age of AI

Modern AI applications do not operate in isolation, they ingest massive volumes of data, rely on complex open source dependencies, and deploy across hybrid cloud environments.

AI introduces novel attack vectors, such as:

- **Data poisoning and tampering.** Bad actors may manipulate training datasets or model weights to compromise decision-making.
- **Software supply chain vulnerabilities.** Unvetted open source libraries, models, or container dependencies containing unpatched CVEs.
- **Data leakage and compliance violations.** Sensitive corporate or personal data exposed via unencrypted workloads or loose access controls.

Without a strong security foundation, high-level AI safeguards (such as guardrails or prompt filters) remain vulnerable to underlying system exploits. A resilient foundation supports data privacy, compliance standards, and operational integrity and is enforced from the OS up to the model pipeline. By making sure the foundation is safeguarded by default, the AI models and sensitive data they process are safeguarded from the ground up, reducing the attack surface from the start.

Recommendations and best practices

Reduce software supply chain security risks by using open source software from a trusted enterprise open source vendor, such as Red Hat that provides support throughout the entire lifecycle of their software. An enterprise open source vendor develops their software with a reliable software supply chain security process that includes curating open source software on behalf of their customers. This ensures that the open source software that customers use is trustworthy, resilient, and security-focused.

AI pipelines heavily use open source tools and pretrained components. Red Hat® Product Security mitigates this supply chain risk by directing continuous vulnerability scanning, patching, active software component maintenance, and lifecycle management. Every foundational image is built on a security-focused, SLSA3-compliant build pipeline that is tracked continuously by Red Hat Product Security, delivering rapid testing, building, and delivery of timely and verified, critical security updates to the Red Hat portfolio.

In addition, it is important to run critical applications on top of a platform with built-in security capabilities. This will provide the foundational security from which organizations can reliably run applications and AI workloads. Platform-level and hardware-enabled security represents the literal foundation of any layered security approach, providing the essential initial trust that higher-layer security controls depend on. Without base-level OS and container isolation, kernel-space exploits invalidate upper-tier AI validation layers.

Red Hat provides an enterprise-grade, zero trust security architecture designed to support AI workloads across on-premise, cloud, and edge environments.

Prioritize a security-focused foundation for applications and AI by adopting a resilient, trusted OS hardened for stability and security like Red Hat Enterprise Linux. A trusted OS allows you to reliably scale critical applications and AI workloads, maintain security compliance, and roll out emerging technologies consistently across bare-metal, virtual, container, and all types of cloud environments.

Red Hat Enterprise Linux

As the standardized, enterprise-grade foundation for the Red Hat portfolio, Red Hat Enterprise Linux delivers built-in security features tailored for AI. It protects sensitive training data and proprietary models using confidential computing, which safeguards data in memory during processing—even from privileged administrators or cloud providers—guarding against memory attacks and malicious code.

Red Hat Enterprise Linux can help protect against AI-related threats using:

Strong data protection

With the capability for data encryption at rest, such as Linux unified key setup (LUKS) and network bound disk encryption (NBDE) and data in transit such as transport layer security (TLS) and secure shell protocol (SSH), Red Hat Enterprise Linux helps safeguard the vast amounts of data used by AI models. Red Hat Enterprise Linux 10 is the first enterprise Linux distribution to ship NIST-approved post-quantum algorithms, such as Module-Lattice Key-Encapsulation Mechanism (ML-KEM) and Module-Lattice Digital Signature Algorithm (ML-DSA), by default, helping protect today's AI training data and model weights against tomorrow's harvest now, decrypt later threats.

Security-focused software supply chain

Access all content through Red Hat's dedicated, security-focused content delivery network (CDN), maintaining trusted, consistent, and tamper-proof updates.

Robust access control and isolation

Security-Enhanced Linux (SELinux) enforces mandatory access controls to isolate processes and containers, mitigating CVEs and preventing privilege escalation—even from compromised AI software. Paired with Address Space Layout Randomization (ASLR), enabled by default, Red Hat Enterprise Linux raises the bar against the memory-corruption exploits AI workloads increasingly face. Red Hat Identity Management (IdM) centralizes authentication, enforces multifactor authentication, and closes the over-permissioned accounts and credential sprawl that attackers target first.

Application allowlisting combines file access policy daemon (fapolicyd) with integrity measurement architecture (IMA) and extended verification module (EVM) to make sure only authorized applications execute and files remain untampered.

"Shift left" security and automated defense

Image Mode for Red Hat Enterprise Linux allows organizations to build prehardened images with embedded compliance baselines from the start, while Red Hat system roles automate consistent security at scale. Red Hat Lightspeed proactively scans environments for vulnerabilities and policy drift using AI-driven assistance to recommend and prioritize fixes. Kernel live patching addresses the collapsed time-to-exploit window by applying critical kernel vulnerability fixes with reduced or zero reboot requirements, minimizing disruption to running AI workloads.

By integrating these robust security features, Red Hat Enterprise Linux empowers organizations to more confidently adopt and deploy AI technologies while building and maintaining a strong zero trust architecture that can adapt to new and amplified threats.

Red Hat OpenShift

Run your existing and emerging applications on Red Hat OpenShift®—a comprehensive application platform for virtualized, containerized and AI workloads across any cloud: public, private and edge.

Red Hat OpenShift contains capabilities to help protect against various AI-related threats, including:

Zero trust isolation

Role based access control (RBAC) acts as the first line of defense for AI workloads by strictly verifying who is authorized to deploy and manage resources. Once deployed, Security Context Constraints (SCC) enforce least privilege at the node level by restricting kernel capabilities and preventing the container from compromising the underlying host. Finally, network microsegmentation implements a strict default-deny architecture, trapping any potential breaches and preventing lateral movements or unauthorized data exfiltration.

Continuous runtime protection

Incorporates security into the build, deploy, and runtime workflows with consistent vulnerability management and policy guardrails. A Kubernetes-native security approach provides accelerated resolutions and greater communication between developers and platform teams.

Data sovereignty

Organizations can train and infer models wherever data resides: on-premise, in sovereign public clouds, or at the edge. This maintains strict alignment with regional privacy laws.

By combining continuous supply chain validation, hardened Linux kernel security, and zero trust container orchestration, Red Hat helps enterprises innovate with AI confidently while maintaining control over risk, privacy, and compliance.

Tactical steps

Take these actions when getting started with building a strong security foundation in the age of AI:

1. **Switch to commercially available versions**
   Migrate your open source software directly from upstream open source projects to trusted, commercially available versions. These versions are tested and validated to reduce the risk of bugs and security vulnerabilities. They may also include enterprise support that can quickly deliver security patches and provide guidance on configuring your software for security. By adopting open source software from a trusted enterprise open source vendor, teams can make sure that their software is developed with a robust software supply chain security process and enterprise support is provided throughout the entire lifecycle of the software. This allows enterprises to use open source software while minimizing security risks.

2. **Choose a platform with built-in security features**
   It's important to choose a platform (such as an OS, container application platform, and automation platform) with built-in security capabilities. This will provide the foundational security from which organizations can reliably run critical applications and AI, include multilayered security capabilities to reduce risk, and implement security and compliance automation at scale.

3. **Implement security throughout the technology stack**
   Once a foundational base for security is established, make sure that the layered technologies running on top of that foundation inherit the security benefits and work in tandem for multilayer security.

"We chose Red Hat OpenShift to meet HIN's security and multitenancy requirements. The solution provided the right foundation to build out the appropriate security architecture."

Aarno Aukia, Co-founder, VSHN

Read the Health Info Net case study.

Security pillar 2: Protecting the pipeline (Securing the software and AI supply chains)

AI-powered agents are becoming active participants in the software supply chain—writing code, reviewing pull requests, selecting dependencies, managing deployments, and calling tools autonomously. As their access and autonomy increase, the attack surface can expand faster than traditional controls can adapt.

Traditional supply chain risks, such as compromised dependencies, unsigned artifacts, and positioned continuous integration and continuous deployment (CI/CD) now intersect with novel ones: prompt injection, unvetted model weights, hallucinated packages, and agents with excessive permissions.

As software supply chains expand to include not just code, but models, datasets, and increasingly large artifacts, the question is no longer whether open software ecosystems can scale, but who pays for that scale, and how long the current system can hold. A single compromise in any component propagates vulnerabilities downstream, threatening data privacy, regulatory compliance, and operational integrity. In 2025, software supply chain risk evolved dramatically. Since 2019, over 1.2 million open source malware packages were logged by Sonatype.[5]

Recommendations and best practices

When an AI agent is a first-class actor in the supply chain, practical safeguards need to be in place to protect against these new threats. Key safeguards include provenance and attestation for both code and models, least-privilege agent permissions, human-in-the-loop gates for high-risk actions, and curated allowlists of trusted libraries that keep agents from pulling in unvetted or hallucinated packages.

By using enterprise-grade open source offerings, such as those offered by Red Hat, organizations can take advantage of the more than 30 years of experience Red Hat has in securing the open source software supply chain of their products. In addition, enterprises need solutions designed to lead with AI-powered development and strengthened software supply chain security.

Red Hat OpenShift Platform Plus is a unified platform that includes Red Hat OpenShift, Red Hat Advanced Cluster Security for Kubernetes, Red Hat Advanced Cluster Management for Kubernetes, Red Hat Quay, and Red Hat OpenShift Data Foundation. This platform helps enterprises build, modernize, and deploy containerized applications in Kubernetes securely and at scale. Multicluster security, compliance, application and data management are provided for consistency throughout the software supply chain.

5 Cox, Brian."2026 State of Software Supply Chain." Sonatype, 31 July 2026.

Red Hat Advanced Developer Suite accelerates AI-powered development with solutions that extend the capabilities of Red Hat OpenShift to optimize the software lifecycle. Red Hat Advanced Developer Suite empowers platform engineers to deliver a superior developer experience, strengthen software supply chain security, and accelerate the path from AI experimentation to production-ready intelligent applications. It is a curated bundle of premium developer technologies and services. The suite includes:

- **Red Hat trusted software factory.** Powered by Konflux, it provides a standardized, multitenant build system that mirrors Red Hat's internal security and build standards. Using a policy-as-code approach embedded directly into CI/CD pipelines, it helps set and enforce compliance policies. It uses a centralized policy engine and enterprise contracts to continuously validate builds against security standards, approved images, and cryptographic signatures before any artifact reaches production.
- **Red Hat Developer Hub.** An internal developer portal (IDP) based on Backstage that provides golden paths for accelerated software delivery and AI project initiation.
- **Red Hat Trusted Profile Analyzer.** A centralized console for real-time visibility into software risk, software bill of materials (SBOM) management, and compliance tracking. In addition, exploit-aware security intelligence is integrated to provide AI-powered analysis to prioritize vulnerabilities based on real-world exploitability, helping teams focus on the 1% of risks that actually matter, avoiding "alert fatigue."
- **Red Hat Trusted Artifact Signer.** A service for cryptographically signing and verifying software artifacts, AI models, and metadata to ensure end-to-end provenance.

Lightwell helps enterprises by providing trusted content and ensuring a strong security foundation with remediations for eligible application-layer open source dependencies without forcing disruptive version upgrades. For eligible third-party open source components, Red Hat provides backported security fixes for the library versions organizations already run in production, along with supporting security artifacts as defined in the applicable offering terms and conditions. Teams access Lightwell remediations through Lightwell repositories and can integrate them into existing build processes alongside the public open source repositories already in use.

Red Hat Hardened Images offers a free, vendor-neutral catalog of trusted, micro-sized images created from Red Hat's years of expertise. These images provide a foundational set of core content, including languages, runtimes, databases, web technologies, and developer tools, all designed to reduce the burden caused by CVEs. By reducing the images' attack surface and moving security to the start of the build process, Red Hat Hardened Images reduces toil and lets teams stop chasing scanner noise and start shipping code. Red Hat Hardened Images can run on any Linux platform.

Tactical steps

Adopt these actions when securing the software and AI supply chains:

1. **Start small and expand**
   Choose a single project to begin. Encourage experimentation and iterative, continuous improvement to fine-tune and optimize your process. Celebrate successes and showcase proven value to others within your organization.

2. **Set clear, agreed-upon goals and timelines**
   Transparency is key. Everyone involved needs to understand and agree with the goals and timelines for the project.

3. **Cross-train your staff**
   Establish learning paths about security, infrastructure, AI, and development that are regularly updated and readily available to all team members.

4. **Create a security work group**
   Build an integrated, cross-discipline team to define security use cases and strategies. Learn from other organizations.

5. **Implement security across the supply chain pipeline with a unified application platform**
   Security needs to be implemented across the software development life cycle (SDLC) using a trusted, unified application platform, hardened for security. Securing the software supply chain when incorporating autonomous AI agents into the SDLC requires shifting from static scanning to active runtime governance.

These are key tactical steps to implement security across the software and AI supply chain pipeline:

Enforce ephemeral nonhuman identities (NHI). Assign AI agents unique service accounts with short-lived, task-scoped OpenID Connect (OIDC) tokens rather than static credentials. Restrict permissions to feature branches using fine-grained role-based access control (RBAC)—never give agents access to main branches or production secrets.

Isolate execution in air-gapped sandboxes. Run agent tasks in disposable containers or MicroVMs (e.g., Firecracker) with default-deny egress rules. Allow outgoing network traffic only to approved LLM endpoints, internal source control, and pre-approved package registries.

Route package downloads through scanned proxies. Point agent configuration files (pip.conf, .npmrc) exclusively to an internal package repository (e.g., Red Hat Quay, Nexus). Block requests for unverified external libraries to protect against slopsquatting and package hallucination attacks.

Sanitize context stream against prompt injection. Preprocess external tickets, pull request comments, and web context through input-filtering pipelines before passing them to the agent. Separate system control instructions from raw user data using strict structural delimiters to block indirect prompt injection.

Mandate AI Software Bill of Materials (AI-SBOMs) and human approval. Attach an AI-SBOM to all agent-generated pull requests (PRs), recording the model, prompt hash, and dependencies used. Enforce automated static application security testing (SAST) and software composition analysis (SCA) security scans, requiring mandatory human review and cryptographic signing before any code merges into production pipelines.

Bank of India builds governance into application delivery
To strengthen application delivery while maintaining strict compliance, Bank of India used Red Hat OpenShift Platform Plus to build security directly into its development lifecycle. By codifying security checks, quality gates, and regulatory controls directly into its CI/CD pipelines, the bank consistently enforces governance across all clusters without slowing development.

Read the Bank of India case study.

Security pillar 3: Implement security, safety, and governance for agentic AI

The shift from generative AI assistants that respond to prompts to autonomous agents that plan and execute multistep tasks introduces a new category of operational risk. Unlike conversational AI applications, agentic systems can be authorized to access databases, call APIs, modify cloud infrastructure, execute code, process transactions, and retrieve external content.

This access turns model errors and AI-specific attacks into potential operational actions. A chatbot that hallucinates may produce an incorrect answer. An agent that misinterprets instructions, or is manipulated through prompt injection, could alter the wrong resource, expose sensitive data, or initiate an unauthorized transaction. The greater the agent's autonomy and access, the more important it becomes to establish clear identity, isolation, authorization, monitoring, and approval controls.

Recommendations and best practices
To combat machine-speed threats, organizations must move beyond traditional perimeter defenses and implement robust security, safety, and governance practices tailored to AI.

- **Assume breach with defense-in-depth.** Start with the assumption that an agent will be compromised. Implement a layered defense strategy where each layer assumes the one above it has failed to contain the threat
- **Fortify nonhuman identities (zero trust).** With nonhuman identities multiplying in the enterprise, replace static, shared API keys with cryptographically verifiable, short-lived workload identities for every agent.
- **Sandbox code execution.** Never run agent-generated code in unrestricted environments. Decouple the agent's reasoning from its execution by using ephemeral, kernel-enforced sandboxes with strict file system and per-binary network restrictions.
- **Connect security and governance.** Avoid dispersed AI tools. Establish strict approval processes and centralized governance gateways to track which agents have access to which tools, actively monitoring for shadow AI.
- **Apply proactive red teaming and guardrails.** Do not wait for deployment to find flaws. Use automated red teaming to simulate adversarial attacks (such as jailbreaks) in CI/CD pipelines, and deploy inline guardrails to filter unsafe inputs and outputs in real-time.

How Red Hat AI helps

Red Hat delivers a unified, 6-layer MLSecOps architecture to boost security focus for the agent lifecycle from the hardware to the model:

- **A trusted foundation.** Security starts at the OS and infrastructure level. Red Hat Enterprise Linux and Red Hat OpenShift use SELinux, restricted SCCs, and confidential containers to provide hardware-level isolation for AI workloads, keeping them inaccessible to unauthorized host processes.
- **Process-level sandboxing.** Red Hat contributes to and will integrate with **OpenShell**, providing kernel-enforced sandboxing. OpenShell limits blast radiuses through credential-free inference, landlock file system allowlisting, and per-binary network policies (e.g., allowing a node process to reach an API while blocking an injected curl command from exfiltrating data).
- **Agent identity and governance.** Using OpenShell and Secure Production Identity Framework for Everyone (SPIFFE)/SPIFFE Runtime Environment (SPIRE), Red Hat's zero trust workload identity manager assigns cryptographic identities to every agent. Additionally, a model context protocol (MCP) gateway centralizes tool aggregation and enforces rate limits and zero trust authentication on every tool call an agent attempts. MCP gateway also allows tool-level authorization policies, tied to workload identities, providing much more granular control than most MCP servers implement.
- **AI safety and guardrails.** Red Hat OpenShift AI integrates **TrustyAI** and **NVIDIA NeMo Guardrails** to intercept unsafe prompts before they reach the model and validate outputs. For proactive defense, Red Hat integrates **Garak** and technology from **Chatterbox Labs** to continuously red-team models against vulnerabilities like prompt injections and data leakage.

This end-to-end stack makes sure autonomous agents operate securely, maintaining enterprise data sovereignty and minimizing systemic operational risk.

Tactical steps

Use these actions to get started with implementing security, safety, and governance for agentic AI.

1. **Harden the foundation.** Adopt an underlying platform that is security-focused, uses restricted SCCs, SELinux, and confidential computing to prevent unauthorized host access.
2. **Enforce cryptographic identity.** Replace shared API keys with zero trust workload identities (such as SPIFFE/SPIRE) so every agent has a verifiable, short-lived identity.
3. **Sandbox code execution.** Run agent-generated code in isolated, process-level environments (like OpenShell) that enforce per-binary network policies, filesystem allowlisting, and credential-free inference.
4. **Deploy guardrails and centralized gateways.** Use tools like NeMo Guardrails to intercept prompt injections, and an MCP Gateway to centrally authenticate, rate-limit, and govern every tool call an agent attempts.
5. **Red team and automate response.** Continuously probe models and guardrails for vulnerabilities using automated adversarial testing (e.g., Garak), and use the results of this testing to update your guardrail configurations.

How ARSAT provided more responsive services and better connectivity with Red Hat OpenShift AI

Argentine satellite operator ARSAT needed to integrate and automate complex supply network operations without introducing risk to critical infrastructure. By standardizing on Red Hat OpenShift AI, ARSAT established a security-focused, centralized foundation for deploying automated intelligence. The platform enables automated end-to-end supply chain processing with built-in platform security.

Read the ARSAT case study.

Security pillar 4: Automating security at scale for machine-speed defense

AI has fundamentally accelerated the cybersecurity threat landscape, empowering attackers to autonomously discover and chain vulnerabilities to build functional exploits in seconds. Consequently, organizations face a critical mismatch: the time-to-exploit window has collapsed, yet traditional manual incident responses operate on a human-scale timeline.

Defending against these machine-speed attacks requires automating security at scale to eliminate human latency. Relying on manual triaging and deployment causes security teams to become overwhelmed by alert fatigue and operational bottlenecks, leaving systems dangerously exposed to rapid exploitation.

Automated security enforcement allows organizations to respond at the true velocity of AI threats. By using automation, security and IT operations teams can evaluate alerts, instantly isolate compromised systems, revoke credentials, and deploy targeted controls. This shrinks the detection-to-containment cycle from days down to seconds, effectively breaking the AI exploit chain before lateral movement can occur.

Ultimately, human patching cycles cannot keep pace with recursive AI exploitation engines. Automating security at scale is essential to transform reactive defenses into a continuous, machine-speed response that builds a resilient, self-defending enterprise.

Recommendations and best practices

AI models such as Mythos can autonomously discover and weaponize vulnerabilities in hours, compressing the time-to-exploit window and rendering traditional human-scale patching cycles obsolete. To defend against these machine-speed attacks, organizations must move beyond manual incident response and automate security at scale.

- **Transition to event-driven defense.** Shift away from rigid, scheduled patching and maintenance windows, which create predictable vulnerabilities. Instead, implement continuous, event-driven enforcement to instantly contain threats the moment they are detected.
- **Enforce zero trust principles.** Adopt a defense-in-depth and zero trust architecture that assumes breaches are inevitable. Establish an automated operational layer where every deployment, patch, or configuration change is strictly gated by policy.

- **Use dynamic credentials and micro-segmentation.** Shrink the attack surface by replacing persistent secrets with short-lived, dynamic credentials, and enforce strict network access controls to prevent lateral movement of threats.
- **Automate triage to reduce alert fatigue.** Utilize automation to evaluate incoming security alerts against predefined rules, helping security teams instantly filter out false positives and focus only on genuine high-risk threats.
- **Automate risk mitigation.** Replace subjective Common Vulnerability Security Scoring (CVSS) with deterministic control-aware environment scoring. Verify, remediate and quantify hardening posture across your fleet in a single workflow.
- **Patch to production.** Lightwell remediates vulnerable open source dependencies while Red Hat Ansible® Automation Platform enforces controls and triggers patching workflows for your critical applications.
- **Compliance as code.** Enforce your security baselines as code, detect drift the moment it occurs and produce audit ready reporting and remediation.
- **Enforce at scale.** Move security policies from static documents to living automation controls. Detect violations before and during automation workflows.

Red Hat Ansible Automation Platform provides the necessary speed, scale, and intelligence to close the critical gap between threat discovery and remediation.

- **Policy enforcement point (PEP).** Ansible Automation Platform sits at the center of an organization's zero trust architecture, ensuring every operational action passes through identity checks, policy evaluation (via Open Policy Agent), and audit logging before execution. This centralized enforcement breaks the AI exploit chain by blocking unauthorized operational paths across the enterprise.
- **Machine-speed containment.** Using Event-Driven Ansible, the platform processes alerts from security tools, such as security information and event management (SIEM), to instantly trigger predefined workflows. It can automatically isolate compromised networks, revoke active database credentials, and update firewall rules in seconds. This shrinks the detection-to-containment loop to match the velocity of AI-driven attacks, eliminating critical human latency.
- **Operational uptime protection.** If an automated containment action or configuration change disrupts production, Ansible Automation Platform can automatically trigger rollbacks while instantly escalating the alert to human engineers. This human-in-the-loop fallback maintains business resilience without leaving systems exposed to the initial threat.

Tactical steps

Use these actions to get started with security automation.

1. **Begin with a single project.** Do not try to automate everything at once. Choose a limited set of tasks to start with.
2. **Choose repetitive tasks.** Automate tasks that are performed repetitively, including configuration management, software package and patch management, security vulnerability identification and remediation, and policy enforcement.
3. **Automate patching at scale.** Transition from manual maintenance windows to continuous, orchestrated patching across the infrastructure to quickly close vulnerability gaps.
4. **Automate threat containment.** Deploy automated workflows that trigger instantly upon threat detection (such as a brute-force attack) to isolate compromised networks or revoke database access, shrinking response times to seconds.
5. **Implement event-driven triage.** Use event-driven automation to instantly evaluate incoming security alerts against predefined rules, filtering out false positives so teams can focus on genuine threats
6. **Enforce policy as code.** Establish a central policy enforcement point (PEP) to ensure every deployment or configuration change is strictly gated by automated identity and policy checks.
7. **Deploy dynamic credentials.** Shrink the attack surface by using automation to replace persistent secrets with short-lived, dynamic credentials that expire immediately after a task is complete.

Kreditplus modernizes services with Red Hat technologies
Financial institution KreditPlus replaced complex legacy maintenance routines with a unified Red Hat footprint (Red Hat Enterprise Linux, Red Hat OpenShift, and Ansible Automation Platform), empowering their security teams to keep infrastructure up to date with the click of a button while providing developers with compliant self-service environments.

Read the Kreditplus case study.

Ready to get started?

Security for an AI-driven world needs to be continuous and at every layer. No matter where an organization is along its AI journey, Red Hat can help.

Red Hat embeds a security focus at every layer across its products, services, and support using open source security principles to help organizations protect infrastructure, applications, and AI while maintaining full data control, sovereignty, and compliance.

As a leader in enterprise open source infrastructure, application development, AI, and automation solutions, Red Hat provides trusted open source software that helps organizations implement security, in a layered approach, from the OS to AI agents, to better safeguard workloads on premise, in the cloud, or at the edge, in both connected and disconnected environments. Red Hat also delivers enterprise-grade support, hands-on training, and expert services to help teams build and operate hybrid cloud environments in the age of AI more efficiently and with a greater focus on security.

Explore these resources to strengthen security across your infrastructure, software and AI supply chains, and AI agents:

- Understand AI security and its role across the AI lifecycle.
- Explore strategies for defending against AI-accelerated threats.
- Learn how Red Hat Services can support your security initiatives through expert-led guidance, hands-on skills development, and proactive support.
- Watch Red Hat Security Symposium 2026 on demand.

Discover Red Hat's approach to security in an AI-driven world.

About Lucy Huh Kerner, Director, Security Global Strategy and Evangelism, Red Hat

Lucy Huh Kerner leads Red Hat's security thought leadership, strategy, and evangelism, encompassing related areas like digital sovereignty and AI. Additionally, she creates and presents security (and related) technical content to the field, customers, partners, analysts, and press and has spoken at numerous events, including security conferences. Lucy has more than 20 years of professional experience as both a software and hardware development engineer, solutions architect, and global security strategist and evangelist, where she worked on various aspects of security.

About Red Hat

Red Hat is the world's leading provider of enterprise open source software solutions, using a community-powered approach to deliver reliable and high-performing Linux, hybrid cloud, container, and Kubernetes technologies. Red Hat helps customers develop cloud-native applications, integrate existing and new IT applications, and automate and manage complex environments. A trusted adviser to the Fortune 500, Red Hat provides award-winning support, training, and consulting services that bring the benefits of open innovation to any industry. Red Hat is a connective hub in a global network of enterprises, partners, and communities, helping organizations grow, transform, and prepare for the digital future.

North America	Europe, Middle East, and Africa	Asia Pacific	Latin America
1 888 REDHAT1 www.redhat.com	00800 7334 2835 europe@redhat.com	+65 6490 4200 apac@redhat.com	+54 11 4329 7300 info-latam@redhat.com

facebook.com/redhat
x.com/RedHat
linkedin.com/company/red-hat

redhat.com

Copyright © 2026 Red Hat, LLC. Red Hat, the Red Hat logo, and OpenShift are trademarks or registered trademarks of Red Hat, LLC or its subsidiaries in the United States and other countries. Linux® is the registered trademark of Linus Torvalds in the U.S. and other countries. All other trademarks are the property of their respective owners.