Red Hat

Unlock the full potential of **AIOps with automation**

Align automation, infrastructure, and AI at scale

From AI excitement to enterprise action

AI is no longer a futuristic ambition—it's a present day priority. Across every industry, organizations are striving to capitalize on the potential of AI to improve decision making, automate repetitive work, personalize experiences, and foster innovation. But while the enthusiasm is growing, readiness is not.

In fact, 97% of global executives say they're excited about AI and plan to incorporate it into their operations, yet only 2% feel their organizations are prepared to deploy AI effectively.[1]

What's holding organizations back? Scaling AI for enterprise presents many challenges, but 3 key considerations require attention from all organizations:

1. Infrastructure readiness

Is your IT infrastructure (including networks, storage, data, monitoring, load balancers, and applications) fully connected and optimized for AI workloads? Is it automated and resilient for scaling of AI workflows, while integrating all tools?

2. Fast response to alerts and conditions

Once your AI and observability tools trigger events, do you have automated response and remediation in place to resolve them?

3. Confidence in acting on AI

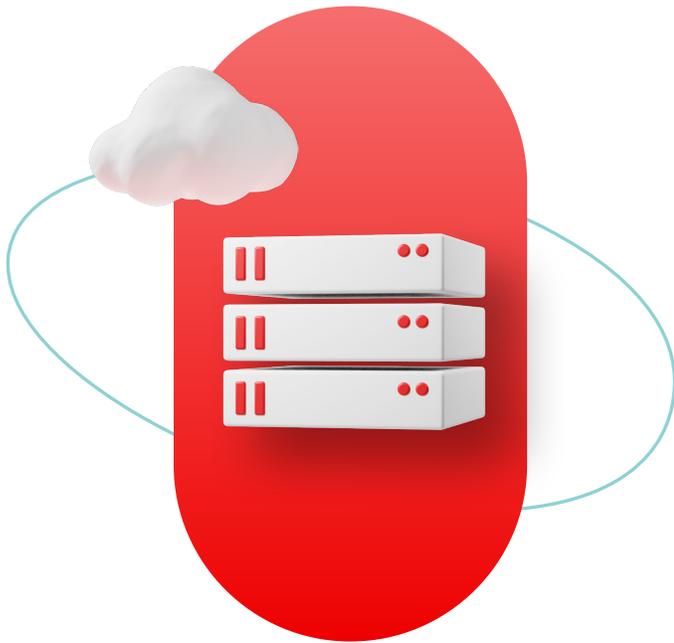Are you confident in your ability to control the automation that's running in your environment?

97%

of global executives say they're excited about AI and plan to incorporate it into their operations.[1]

2%

feel their organizations are prepared to deploy AI effectively.[1]

Answering these questions can help assess how prepared your organization is to scale AI. It can also determine if you have the foundational technology and governance in place to support your AI projects and the teams that build and manage them at scale.

---

[1] "Cisco's 2025 AI Briefing: CEO Edition." *Cisco*, accessed 2 Jun. 2025.
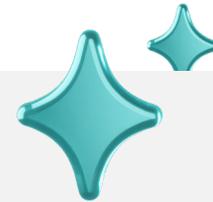
The unseen cost of AI

Unlike traditional workloads, AI platforms require powerful compute infrastructure, specialized hardware, massive data pipelines, consistent environments, and tight controls for compliance and governance. And unlike a static application environment, AI workloads evolve rapidly as models are retrained, tuned, and redeployed.

IT operations teams must spend a growing amount of time diagnosing and manually responding to automated report errors and performance issues.

These new and emerging requirements create layers of operational burden that, while manageable with the right approach and safeguards, are best prepared for early in the AI process.

The predicament that many organizations find themselves in is that AI initiatives were typically championed by data scientists or AI developers. With the introduction of gen AI, many AI initiatives are now led by the business directly. This shifts the focus to innovation experimentation and business outcomes— not day-to-day operations.

Gen AI

But as projects move from the lab to production, they collide with real-world requirements such as availability, scale, cost efficiency, and security. This is where IT operations and automation teams come in–but often they're brought in too late.

IT automation creates opportunity for AI

If AI is to be operationalized successfully, IT automation must be integrated from the start. But many IT Ops leaders, development operations (DevOps) engineers, and site reliability engineers (SREs) find themselves on the sidelines of AI conversations. This is because they're often seen as the team that makes it work after an AI solution is defined. For many organizations, this is a missed opportunity.

Automation isn't just a supporting tool for AI; it's a way to amplify its potential. In fact, 69% of CEOs say AI will improve operational efficiency and reduce costs, however, 35% are concerned about the limitations of their infrastructure. More than half (55%) of CEOs feel they need to upgrade their network infrastructure and operations to get ready for AI.[1]

Some of the biggest concerns that IT leaders have about their AI infrastructure include:[1]

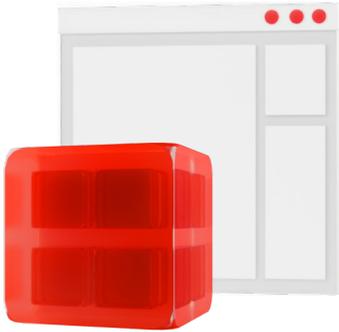|			
Not having the right infrastructure to support AI workloads.	A lack of clear policies around AI safety and governance.	Excessive time spent manually addressing performance issues.	Limited visibility and automation across hybrid environments.

[1] "Cisco's 2025 AI Briefing: CEO Edition." *Cisco*, accessed 2 Jun. 2024.

IT infrastructure meets AI challenges

As AI moves from pilot projects to business-critical deployments, it introduces new challenges that require organizations to reassess the capabilities of their traditional IT environments.

Infrastructure readiness has declined globally even as organizations anticipate the increase in AI workloads. The *Cisco 2024 AI Readiness Index* measured companies across 6 key pillars: strategy, infrastructure, data, governance, talent, and culture. The report found that only 15% of organizations have a strong AI infrastructure despite 59% admitting they have a maximum of 1 year to deploy an AI strategy, or they run the risk of negative business effects.[2]

15%

of organizations have
a strong AI infrastructure.[2]

59%

say they have a maximum of 1 year to deploy an AI strategy, or they run the risk of negative business effects.[2]

Understanding the AI readiness gap

While many businesses are investing heavily in model development and AI talent, the gap that is being created (and widening) occurs when organizations overlook the foundational layer required to deploy and scale AI models.

[2] "Cisco AI Readiness Index." *Cisco*, accessed 10 May 2024.

Contributing factors behind the growing gap in AI readiness include:

AI infrastructure complexity

This often includes a mix of bare metal servers, GPUs, networking, high-performance storage, hybrid cloud environments, and edge deployments—all of which add up to a much more complex environment to manage and maintain.

AI is costly and resource-intensive

Managing and training workloads, inference services, and data pipelines demands constant provisioning, monitoring, and tuning of the resources.

AI is dynamic and unpredictable

AI workloads fluctuate dramatically in response usage, requiring more agile scaling.

In addition to these factors is the dual responsibility for infrastructure teams to maintain existing (and aging) applications while simultaneously standing up modern AI services. It's not a simple lift-and-shift strategy. Most organizations must support a hybrid environment that includes:

Virtual machines (VMs) and aging platforms that can't be deprecated due to their criticality.

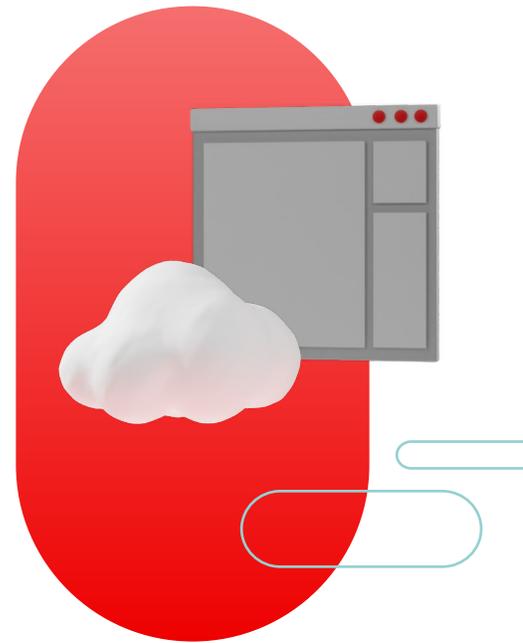Kubernetes clusters hosting containerized AI pipelines.

Cloud and on-premise GPU clusters for training and inferencing.

Complex data flows across systems with varied compliance requirements.

Managing this level of complexity manually or with fragmented toolchains creates bottlenecks, increases risk, and diverts teams from higher-value innovative projects.

Alert fatigue and manual intervention slows progress

For many IT Ops teams, the volume of alerts generated by applications, observability platforms, and security systems has reached unmanageable levels. There is a common misconception that AI is automation. But in practice, teams are still spending too much time investigating automated error reports, diagnosing performance degradations, and coordinating remediation efforts across tools and teams.

These obstacles slow mean time to resolution (MTTR), increasing the likelihood of downtime, and erode the confidence of stakeholders relying on AI to generate outcomes.

Event-driven automation can help close the gap

When specific alerts can trigger predefined, policy-based actions—such as restarting services, updating tickets, scaling resources, or adjusting configuration parameters—organizations gain the agility they need to manage AI workloads in real time with speed, consistency, and accuracy.

Learn more about event-driven automation →

Set safeguards for trusted AI

Effective AI infrastructure goes beyond uptime and efficiency. AI systems increasingly inform and make decisions that can positively or negatively affect customer trust, reputation, and regulatory compliance. IT teams must define and enforce policies to make sure:

AI workloads operate within preapproved usage boundaries.

Data is not sent to unauthorized locations or services.

Changes and updates happen only during defined maintenance windows.

Security and compliance rules are consistently applied.

Aligning to security and compliance standards is important when it comes to automation across sprawling, dynamic environments. Repeatability, visibility, and control are key requirements. Regulatory standards such as the Health Insurance Portability and Accountability Act (HIPAA), General Data Protection Regulation (GDPR), Payment Card Industry Data Security Standard (PCI DSS) Sarbanes-Oxley Act (SOX), and others further complicate infrastructure requirements—especially when AI systems process sensitive or regulated data.

The ongoing process of maintaining compliance can be drastically improved using automation. **Red Hat® Ansible® Automation Platform** allows organizations to codify compliance checks, enforce configuration baselines, and automatically remediate technology drift. This makes sure that the infrastructure supporting AI stays aligned with evolving regulatory requirements.

Bridge the skills gap with automation and AI

AI has contributed to a massive advancement in innovation, but it has also magnified the technical skills gap that many organizations face. As infrastructure becomes more complex and the demands of AI workloads increase, many teams are struggling to find or develop the talent needed to keep up.

Even seasoned infrastructure teams often find themselves at a disadvantage when asked to operationalize AI. This is because they're expected to support hybrid environments, manage GPUs, large volumes of data, and training clusters, enforce policies, and keep services running at scale—all while learning new tools and platforms on the fly.

Between the skills gap and ever-growing portfolio that teams are taking on, operational performance, governance, and innovation are all at risk.

IT automation is key to addressing these challenges

As AI platforms become more entrenched in business-critical workflows, IT teams need ways to deliver repeatable, controlled, and observable operations across every layer of the stack. This means moving beyond traditional runbooks, ticket-based responses, and error-prone manual processes. Automation is the key to bringing infrastructure and AI into alignment.

AI solutions can only be as effective as the infrastructure they run on. Without consistent environments, reliable uptime, and efficient scaling, even the most advanced models will struggle to deliver results. Automation amplifies the effectiveness of AI, especially when integrated with observability and policy engines needed to deploy AI at scale.

Automation helps teams scale AI by:

- Connecting infrastructure components, including networks, storage, observability tools, data repositories and databases, load balancers, and AI platforms, to create a cohesive, orchestrated system.

- Creating consistent configurations so models can be trained and operate reliably across environments.

- Accelerating provisioning of environments required for training, tuning, and inferencing.

- Reducing configuration drift and the risk of performance instability caused by ad hoc changes.

These capabilities are especially important for supporting hybrid or multicloud environments where workloads span on-premise datacenters, private and public clouds, and edge locations.

Safeguards, governance, and control

AI may recommend automating a task or workflow, but that doesn't mean the recommendation aligns with your internal standard operating procedure. As infrastructure scales and automation expands, organizations need to enforce policies to make sure AI systems (and the teams that manage them) perform in predictable, compliant, and security-focused ways.

Using policy-based automation, IT teams can:

Define rules for when and how AI-related services can scale or be modified.	Enforce change windows and access controls.	Limit automation to approved actions or environments.	Make sure infrastructure changes align with compliance and audit requirements.

Automation enhances governance, which is especially important in sectors where AI is being deployed in a regulated environment such as finance, healthcare, and the public sector. For these industries, privacy, accountability, and auditability are non-negotiable.

Turn intelligence into action using automation

AI has an incredible ability to gather and distill information to provide insights, but insights alone are not enough. For organizations to realize real business value from AI, it must trigger outcomes. And to do so, automation is essential.

By integrating automation into AI workflows, organizations can close the loop between detection and response, insight, and execution. Using AI to inform actions, and automation to execute them instantly, accurately, and at scale, is at the core of artificial intelligence for IT operations (AIOps).

Standardize AI infrastructure and operations

Supporting AI workloads demands infrastructure that's scalable, reproducible, and consistent. From standing up Kubernetes clusters for training pipelines to managing GPU usage across hybrid environments, infrastructure teams need to move quickly without compromising stability.

Automation tools, such as those integrated with AI, can alleviate these bottlenecks by optimizing processes and reducing manual workloads.

IT automation plays an important role in orchestrating AI infrastructure by:

Simplifying AI infrastructure deployment across on-premise, cloud, and edge environments.

Managing lifecycle consistency for environments used in training and inference.

Maintaining repeatability so models train in clean, stable environments each time.

Integrating AI workloads into existing IT environments requires coordination between multiple teams, including data scientists, developers, lines of business, and IT operations. By establishing standardized processes and utilizing Infrastructure as Code (IaC) practices, teams can collaborate and streamline the deployment of AI solutions across hybrid and multicloud environments.

Unlock AIOps

The next step forward in operations maturity is going from automated alerts to automated resolution. AI observability tools such as **Splunk**, **Dynatrace**, **Datadog**, or custom machine learning (ML)-based systems can detect anomalies in system behavior. But they still need a downstream mechanism to act on those insights.

Using event-driven automation, you can:

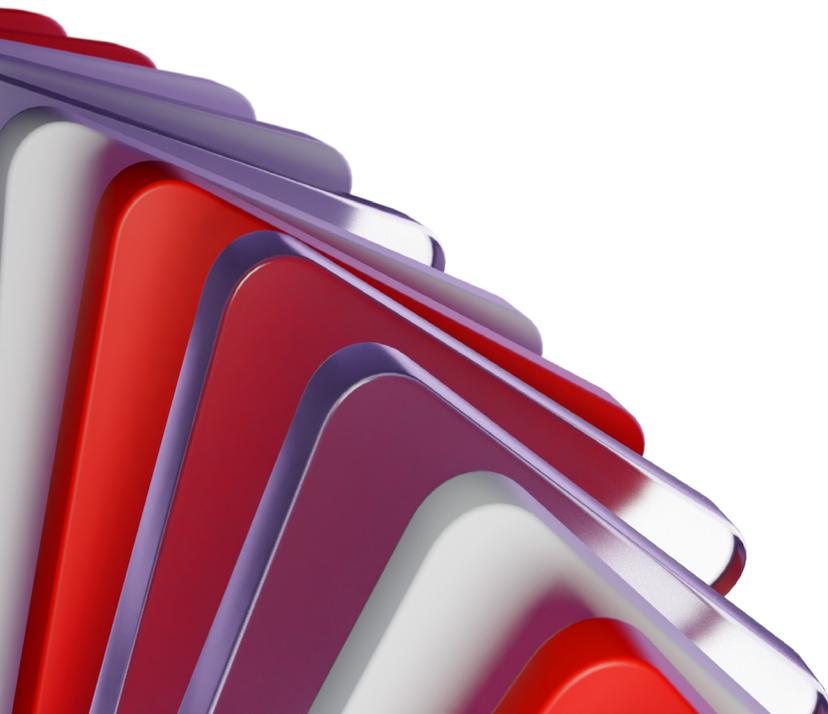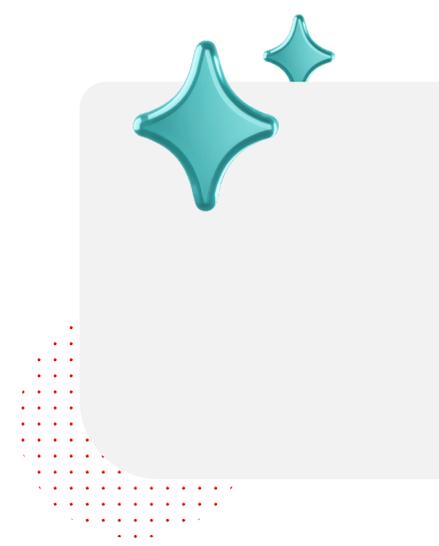Automatically trigger remediation actions based on AI-detected conditions.

Resolve issues more quickly by codifying known responses to recurring incidents.

Integrate AI tools as decision engines while your automation platform becomes the execution plane.

This creates a closed-loop response system where **infrastructure can self-hea**l based on AI-generated insights, reducing MTTR and freeing teams from repetitive tasks.

Execute AI inferences with confidence

As AI models accelerate decisions across customer service, fraud detection, supply chains, and more, safeguards are essential.

Automation helps establish and maintain trust and oversight in AI operations by:

Enforcing compliance rules to data access, usage, and model deployment.

Orchestrating security-focused pipelines that meet internal governance requirements.
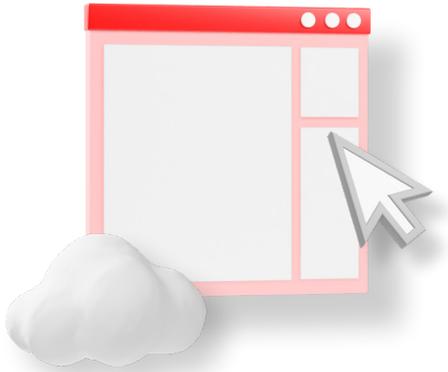
Preventing policy violations such as deploying updates during production hours or spinning up excessive cloud resources.

By embedding these controls into end-to-end processes, organizations can confidently use AI capabilities while maintaining trust and compliance.

Red Hat's strategic approach to IT automation and AI

Successful AI isn't just about building better models; it's about building better systems to support them.

Red Hat's approach to automation and AI is grounded in open source principles, designed to help organizations bridge the gap between innovation and execution. The aim is to empower teams of all skill levels to confidently support AI initiatives, respond to change, and maintain compliance in an increasingly complex IT landscape.

IT automation designed for scale and skill diversity

IT automation removes the complexity of AI infrastructure. With Red Hat Ansible Automation Platform, AI infrastructure is just another domain (in addition to networking, cloud, storage, and virtualization, among others). Ansible Automation Platform provides a consistent way to manage provisioning, configuration, and compliance across AI infrastructure—from bare metal GPUs to cloud environments and edge clusters.

Standardized deployment

Ansible Playbooks provide a consistent and repeatable method for deploying AI components like operating systems, servers, storage, models, containers, data, and networking resources. Using infrastructure as code, Ansible Automation Platform promotes uniformity and reliability across all AI environments, reducing the likelihood of configuration errors or discrepancies.

Data management

One of the most difficult tasks for training AI models is moving the data from where it is created into a location where it can be used for training. Ansible Automation Platform is essential for supporting efficient, large-scale data movement while making sure that the right users have safeguarded and appropriate access to that data.

Close integration with the Red Hat AI portfolio

Ansible Automation Platform offers collections and tooling that make it easy to connect, configure, and coordinate platforms in the **Red Hat AI** portfolio.

To help teams keep up with growing AI demands, Ansible Automation Platform includes features specifically designed to address the growing IT skills gap:

Red Hat Ansible Lightspeed is Ansible Automation Platform's rich gen AI service that enhances user expertise and empowers automation teams to work smarter, make better decisions, and resolve operational issues more quickly. Red Hat Ansible Lightspeed helps boost IT and developer productivity, address skills gaps, and eliminate onboarding challenges. Equipped with Ansible-specific guidance and recommendations, your automation teams can respond to the complex demands of IT.
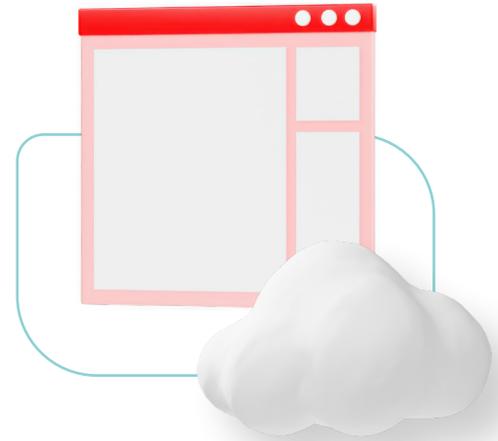
Accelerate automation administration

Simplify Ansible Automation Platform administration with integrated access to documentation while eliminating common points of friction during setup, configuration, and maintenance. Trained on trusted Red Hat data sources, Ansible Lightspeed intelligent assistant helps administrators onboard and troubleshoot issues directly with the Ansible platform experience. It's like having an Ansible expert right at your keyboard.

Create code efficiently

Create Ansible content more accurately and in less time with reliable, trusted code recommendations, served directly in your code editing environment via the Ansible Visual Studio (VS) Code extension. **Ansible Lightspeed** can generate single tasks, multiple tasks, or an entire Ansible Playbook or role from a single prompt. Customize the model using your proprietary Ansible content to further refine gen AI recommendations and improve accuracy.

Maintain code quality

Ansible code bot helps teams keep their automation code bases updated with accepted best practices. **It scans existing content and automatically provides update recommendations** that are ready to review, test, and apply, making it easier to maintain quality and consistency across the development lifecycle. Accelerate Ansible Automation Platform migrations with a simplified way to modernize and harden Ansible code before upgrading to the next version of Ansible Automation Platform.

Red Hat Ansible Lightspeed lowers the barrier to entry into automation and promotes accelerated adoption across infrastructure, DevOps, and platform teams, so automation expertise can scale at the same pace as your organization's AI ambitions.

Respond in real time with Event-Driven Ansible

As AI systems become more operationalized, responsiveness is critical. Whether scaling compute capacity during model training or responding to performance anomalies in production inferencing, infrastructure needs to adapt in real time.

Event-Driven Ansible is a key component of Ansible Automation Platform. It helps teams to define and execute automated actions in response to specific events, such as system alerts, policy violations, or AI-generated insights.

This proactive capability means:

Systems can autoremediate before incidents escalate.

Resource usage can be scaled dynamically.

Policies can be enforced instantly, even in fast-moving environments.

Ticket enrichment can expedite MTTR and service requests.

How does Event-Driven Ansible work?

Event-Driven Ansible connects sources of events with corresponding actions via rules. Ansible Rulebooks define the event source and explain the action to take when the event occurs. Based on the rulebook you design, Event-Driven Ansible recognizes the specified event, checks to see if the conditions you specified are met, then automatically executes the appropriate action. A rulebook can even call your trusted Ansible Playbooks.

By tying observability tools and AI systems directly into Ansible Automation Platform, organizations can move from insight to action much more quickly, fostering AIOps at scale.

AI and automation in action

[Mutua Madrileña](#), a leading Spanish insurer, faced the challenge of managing more than 60 different platforms. This environment was complex, especially following multiple mergers and acquisitions, but crucial to support the company's digital needs. To reduce this complexity, Mutua Madrileña sought to streamline operations and enhance observability.

By integrating Ansible Automation Platform with Dynatrace, Mutua Madrileña implemented event-driven automation to proactively identify and resolve incidents, saving staff time and drastically reducing risk. The result was a 50% reduction in service tickets and improved resolution times.

> By adding an automation stage early in the platform development lifecycle, we've significantly improved velocity rather than this extra step slowing things down.[3]

Marta Ceciliano,
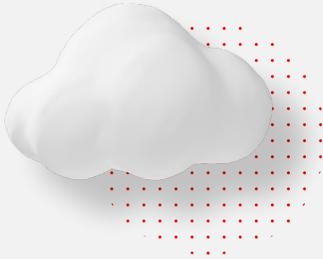Head of Middleware, Automation, and Observability, Mutua Madrileña

Enforce policies to keep AI in check

Enforcing policies at automation runtime is a key best practice for effective AIOps. By defining clear boundaries around what automation can do, you can make sure that AI-powered automation operates strictly within approved parameters—aligning execution with organizational policies and compliance standards. With Ansible Automation Platform, you can **enforce policies** at automation runtime to make sure every AIOps action aligns with intended outcomes.

[3] Red Hat case study. "[Mutua Madrileña adopts Red Hat Automation Platform](#)." 15 Nov. 2023.

Lay the foundation for AI success with automation

Whether you're working to improve infrastructure resilience, orchestrate hybrid environments, or foster intelligent response through AIOps, automation is the critical layer that transforms AI from a strategic vision into real, actionable outcomes.

Take the next step in your AI journey with a consistent platform to collaborate, scale automation, and support your AI ambitions across your organization.

Explore customer use cases

Discover how automation helps orchestrate infrastructure, foster AIOps, and apply policy guardrails to AI.

Read the blog →

Red Hat Ansible Automation Platform and AI in action

Take a closer look at how Red Hat promotes turning AI insights into action.

Explore the use cases →