Manage AI in production with an enterprise platform

Trusted, comprehensive, and consistent

Red Hat AI is a portfolio of products and services that accelerates AI innovation and reduces the operational cost of developing and delivering AI solutions across hybrid cloud environments. Red Hat AI delivers cost-effective solutions with optimized models and efficient inference, simplifies integration with private data, and accelerates delivery of agentic AI on a scalable, flexible platform.

Advancements in AI are transforming business models

Rapid advancements in AI are transforming industries and redefining traditional business models. Enterprise IT organizations are under added pressure to design, build, and deliver AI solutions that provide a competitive advantage. Although AI offers the potential to boost operational efficiency and productivity, many organizations have yet to fully integrate AI technology into their daily operations or realize its full business benefits.

IT organizations must frequently weigh the benefits of using public AI services against the deployment of a dedicated on-premise infrastructure. Data privacy and security remain reminders of the ongoing compromise. While newer government and industry regulations intensify, many technology organizations are navigating this demand by becoming their own Model-as-a-Service (Maas) providers or deploying workloads on sovereign AI cloud environments.

Training and running large AI models at scale can be costly, particularly as organizations deploy multiple models, AI agents, and AI-enabled applications into production. Using domain specific models can significantly refine model responses and predictions. However, customizing these models with business-specific data requires significant compute resources, extensive datasets, and specialized expertise, which can lead to greater expenses. Investments in specialized hardware and training for qualified professionals—to manage these models—can further increase costs and complicate scaling of crucial AI services.

Preparing for the next era of intelligence requires more than adopting current tools. Enterprises must build systems ready to host agentic AI while maintaining the agility to adapt as new models emerge. Without the right platforms and toolset, organizations can lose time trying to piece together multiple technologies, resulting in AI solutions that fail to deliver competitive advantages.

The location of models, AI agents, and data can also influence decisions related to hardware availability, data privacy, and governance. Safeguarding proprietary data and minimizing the costs of complex AI infrastructure can be challenging without the flexibility to deploy AI solutions across distributed environments.

Red Hat offers comprehensive, trusted technologies that speed development and delivery of innovative AI solutions across hybrid cloud environments.

Streamline and speed AI operations

Red Hat® AI is a portfolio of products and services that accelerates time to market and reduces the operational cost of delivering AI solutions across a hybrid cloud landscape. The portfolio supports all stages of an AI adoption journey—from single-server deployments to highly distributed, scalable platform architectures. Designed to simplify AI adoption, Red Hat AI makes advanced technologies more accessible across the entire organization. Using Red Hat solutions help organizations integrate and manage both predictive and generative AI (gen AI) models with increased security at scale. Take advantage of support for a variety of hardware accelerators, original equipment manufacturers

Achieve strategic goals with Red Hat AI

Red Hat AI can be considered for a wide range of use cases:

- Build, migrate, and run ML and predictive AI models.
- Build, deliver, and run gen AI applications.
- Tailor AI solutions with relevant enterprise data.
- Deploy private AI solutions on-site or in air-gapped environments.
- Operationalize and automate model lifecycles via MLOps and DevOps.
- Build multiarchitecture AI deployments.

(OEMs), and cloud providers to provide a stable, optimized, and high-performance environment. IT organizations can deploy critical AI applications and services across diverse environments, including on-site infrastructure and public cloud resources.

The product portfolio for Red Hat AI includes:

- Red Hat AI Enterprise, for enterprises looking to deploy and scale efficiently and anywhere.
- Red Hat AI Inference Server, for optimized inference of large language models (LLMs).
- Red Hat OpenShift® AI, for distributed Kubernetes platform environments.
- Red Hat Enterprise Linux® AI, for individual Linux server environments.

These solutions deliver open source technologies and models, providing access to the latest AI tools curated and integrated across the entire organization. Additionally, the Red Hat AI Partner Ecosystem helps speed the pace of innovation with a range of tested, supported, and validated products and services that address both business and technical challenges.

Increase efficiency with high-performance AI inferencing

Choosing the right model is a critical first step in any AI solution because it defines the system's intelligence. Meanwhile, the efficiency of AI inference is what determines its real-world performance. Red Hat AI helps organizations choose the right model, while allowing them to optimize AI inference across their hybrid cloud—leading to more cost-effective and consistent deployments. The portfolio helps IT teams build predictive models, tune gen AI models, and deploy a combination of both across hybrid cloud environments, empowering enterprises to reach operational efficiency.

Red Hat AI includes key features for increasing efficiency:

- **For predictive AI,** this efficiency is realized through technologies like single and multimodal serving capabilities, which allow enterprises to consolidate hundreds of models onto minimal infrastructure. It also provides support for various optimized runtimes to execute different types of models, such as: virtual large language model (vLLM), OpenVino, NVIDIA NIM, Triton Inference Server, Caikt and, TGIS.
- **For gen AI,** Red Hat AI includes the vLLM runtime to maximize memory use, speed up responses, and increase graphics processing unit (GPU) use for efficient model inference. This allows enterprises to run foundational and custom models—large or small—with significantly lower latency and reduced resource consumption. For organizations requiring even greater scale, the platform integrates large language models distributed (llm-d), a framework that optimizes resource use across the entire cluster. This dual-layer approach allows enterprises to meet strict service level objectives (SLOs) by maximizing the performance of individual models while intelligently balancing workloads across complex, distributed environments.

Red Hat AI also provides access to a collection of optimized and validated third-party models that run efficiently on vLLM across the platform, including performance benchmarks and accuracy evaluations. These models include full model details, SafeTensor weights, and commands for rapidly deploying on Red Hat AI. The platform also allows users to create their own optimized model versions by taking advantage of the model optimization capabilities.

By understanding these high-efficiency inference technologies, IT organizations can move beyond simple implementation to become their own Model-as-a-Service (MaaS) providers. Red Hat AI offers a unified platform that supports data sovereignty by allowing enterprises to serve both predictive and generative models across the hybrid cloud, a flexible deployment strategy. This unified environment provides the reliable foundation needed to power the next generation of agentic applications with both operational agility and regulatory compliance.

Simplify and speed model customization

Gen AI models are typically trained on generic data, which may not provide the specific context that an organization needs for accurate responses and meaningful insights. Red Hat AI helps IT team's AI engineers, data scientists, and domain experts solve unique business challenges—from developing intelligent chatbots and virtual assistants to building sophisticated predictive models for regression and classification. Through a consistent, simplified AI tooling experience, Red Hat AI supports users to build machine learning (ML) models, customize gen AI models, and connect models to enterprise knowledge sources. The portfolio offers a modular architecture for model training, tuning, data ingestion, and synthetic data generation, ensuring that each customization effort is both performant and reproducible.

Red Hat AI includes several key features:

- **Simplified data integration:** Red Hat AI includes data ingestion and preprocessing capabilities allowing IT teams to use structured and unstructured private data for model training and tuning. It also includes tooling for expanding and refining datasets with the synthetic data generation hub.
- **Advanced customization patterns:** Red Hat AI supports a tiered approach for customizing LLMs with private, enterprise data. The platform offers prompt design to enhance genAI model responses and achieve more specific and accurate outcomes. For real-time accuracy, retrieval augmented generation (RAG) allows models to access verifiable enterprise sources, ensuring responses are current and fact-based. For deeper alignment, Red Hat AI provides tooling for model customization that ranges from full fine-tuning to parameter efficient methods with the goal of balancing performance and efficiency.
- **Collaborative, self-service development:** To maintain operational agility, Red Hat AI provides self-service access to popular IDEs and open source frameworks for building predictive models and tuning generative AI. This environment simplifies the allocation of hardware acceleration, ensuring that training jobs are cost-efficient. From the Synthetic Data Generation (SDG) Hub to the training hub, each step of the lifecycle is optimized for transparency and consistent governance across the hybrid cloud.

By providing a simplified and consistent experience, Red Hat AI empowers teams to reduce the gap between generic models and proprietary expertise. This modular approach to connecting models with private data allows for efficient, domain-specific customization that significantly improves the accuracy and relevance of AI responses.

Accelerate agentic AI innovation

While generative models redefined data interaction, the next evolution is agentic AI–autonomous systems capable of reasoning, initiating multistep tasks, and accessing external tools to meet business goals. Red Hat AI provides a flexible, stable foundation designed to simplify this transition, moving organizations from content generation to intelligent, autonomous workflows.

Building production-ready agents requires more than prompting. It demands a unified architecture to coordinate reasoning, orchestrate tools, and govern behavior. Red Hat AI provides the core platform services to provide consistent, repeatable processes across the hybrid cloud while supporting the development, integration, and monitoring of agents at scale.

Red Hat AI includes several key features for accelerating agentic innovation:

- **Standardized integration and orchestration:** Red Hat AI provides a unified application processing interface (API) experience through an enterprise implementation of the Llama Stack API. This standardized entry point simplifies operations via a pluggable architecture that allows agents to use consistent tool calls across different model providers. The process makes certain that as reasoning models evolve, the organization's underlying agentic logic and tool integrations remain stable and portable.
- **Governed tool connectivity:** Through support for the Model Context Protocol (MCP), Red Hat AI provides an open standard for how agents interact with tools, data, and memory. MCP allows agents to discover and invoke tools across APIs and databases reliably to reduce custom integration overhead. To maintain reliable behavior, the AI platform acts like a security backbone by providing core components for guardrailing AI safety and compliance. This protocol includes a roadmap toward an MCP gateway to manage the security complexities and permissions of autonomous tool use.
- **Unified management and experimentation:** Red Hat AI provides a collaborative environment for moving agents from POC to production. This is delivered through 2 consolidated dashboard experiences:

  a. The AI hub, which empowers platform engineers to manage the lifecycle and governance of AI assets.

  b. The gen AI studio, which provides AI engineers a hands-on environment for experimentation and prototyping.

  By addressing the distinct needs of both roles, the platform provides a flexible, unified foundation for building production-ready autonomous workflows.

Red Hat AI empowers enterprises to build autonomous agents that are powerful, predictably intent, and compliant. Organizational IT leaders can now operationalize agentic AI in a scalable and trusted way—from customer support triage to automated IT remediation.

Gain the flexibility to deploy AI solutions anywhere

Red Hat AI provides the flexibility to train, tune, deploy, and run gen AI models and applications wherever it best aligns with business needs. This approach helps meet data privacy, security, and compliance requirements while optimizing hardware infrastructure costs.

With a focus on enterprise AI workloads, Red Hat AI delivers a trusted, consistent, and comprehensive platform for managing AI in production. The platform orchestrates model integration into both new and existing applications, while unifying the management of models, applications, and code into a single location. As a result, deployment and management of predictive and gen AI models will operate across diverse environments with consistency, stability, and flexibility both on site and in the cloud.

Red Hat AI prioritizes security, cost optimization, and operational efficiency to support enterprise AI strategies. It offers optimized inference and serving runtimes like vLLM to increase the efficiency of LLMs at inference time. A range of deployment options across different hardware accelerators, cloud providers, and OEM server environments provides the flexibility needed to balance cloud spend, data storage, and GPU availability.

Effective AI implementation also requires streamlined model lifecycle management. Red Hat AI simplifies this process with reliable machine learning operations (MLOps) and large language model operations (LLMOps) capabilities—including enhanced automation, monitoring, governance, resource allocation, and security. The platform abstracts the complexity of provisioning development environments and managing hardware acceleration for training and tuning, allowing you to focus on AI innovation rather than infrastructure challenges. For organizations with strict data security requirements, Red Hat AI supports on-site and air-gapped deployments, reducing the risk of exposing sensitive data. This level of security makes sure that proprietary data never leaves the organization's control, providing a path to digital sovereignty without sacrificing the speed of the public cloud.

Choose Red Hat AI for every stage of your journey

The true value of enterprise AI is realized at the intersection of powerful models and proprietary business expertise. Red Hat AI provides a modular, integrated portfolio designed to turn AI potential into production reality. This architecture allows organizations to move swiftly from initial pilots to organization-wide deployments, ensuring that each AI initiative is grounded in operational efficiency and measurable business benefits. The Red Hat AI portfolio empowers teams to adapt cutting-edge innovation to their specific data and use cases without sacrificing the control required for enterprise governance.

At the center of this strategy are 2 primary offerings designed to maximize your AI investment. Red Hat AI Inference Server serves as a high-performance engine to optimize, manage, and scale model inference across any footprint from the datacenter to the network edge. Simultaneously, Red Hat AI Enterprise provides a centralized, collaborative environment that unifies the entire AI lifecycle, allowing teams to develop, tune, and deploy models and agentic workflows with speed and precision. Together, these solutions provide the flexibility to choose any model and run it on your ideal combination of hardware and cloud providers, transforming the complexity of the hybrid cloud into a strategic competitive advantage.

The portfolio also includes OpenShift AI for distributed Kubernetes platform environments and Red Hat Enterprise Linux AI for organizations running LLMs in individual Linux server environments.

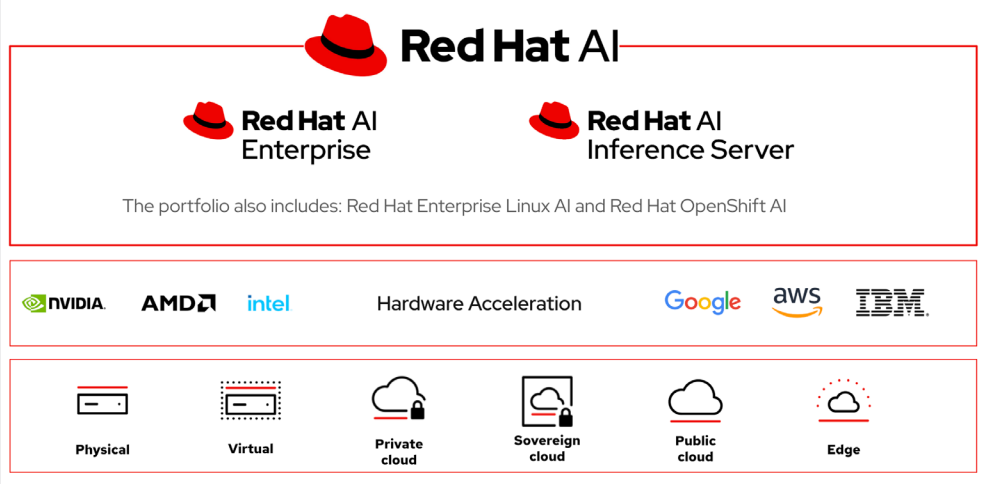

Figure 1. Red Hat AI is a portfolio of products and services that support a range of innovative AI solutions.

Start inferencing with Red Hat AI Inference Server

Red Hat AI Inference Server provides rapid, cost-effective inferencing at scale. With the vLLM inference runtime at its core, the solution gives organizations a unified, high-performance platform to run your IT team's choice of models across hardware accelerators, Kubernetes and Linux environments, and IT infrastructure from on-site datacenters and clouds to edge deployments.

Red Hat AI Inference Server includes access to optimization capabilities including a model repository and LLM compressor—to help models run more efficiently, use fewer resources, and reduce inference costs. The model repository offers a collection of validated and preoptimized models, ensuring rapid deployment with benchmarked performance. The LLM compressor helps reduce the size of models and increase inferencing speeds using advanced quantization techniques, all while maintaining accuracy. Together, these components support accurate and cost-effective inferencing across a wide range of applications.

Deploy and scale anywhere with Red Hat AI Enterprise

Red Hat AI Enterprise is an integrated AI platform for deploying and managing efficient, cost-effective AI models, agents, and AI-powered applications across the hybrid cloud. By unifying the lifecycles of models and applications into a single, ready-to-use development environment, the platform increases operational efficiency and mitigates the risks of AI adoption. With Red Hat OpenShift at its core, Red Hat AI Enterprise ensures business continuity while delivering a consistent experience across bare metal, virtual machines, and public clouds. This centralized

Kubernetes infrastructure allows teams to build modern intelligent applications using familiar tools and frameworks, managing everything from predictive AI development to the scaling of gen AI and autonomous agentic workflows.

The platform empowers IT teams to align foundation models with enterprise-relevant data through advanced RAG and fine-tuning patterns. To boost productivity, Red Hat AI Enterprise provides self-service access to a comprehensive suite of tools—including popular integrated development environments (IDEs) like Jupyter Notebooks, widely used frameworks like PyTorch, and customizable workbenches. A centralized catalog and registry further simplifies innovation by providing access to validated, optimized models and MCP servers.

A unified approach to MLOps, genAIOps, and AgentOps eliminates the complexity of fragmented tooling, allowing teams to move uninterrupted from experimentation to production. To maintain gen AI model performance at scale, Red Hat AI Enterprise includes vLLM to deliver high-throughput, low-latency inference at the engine level, while llm-d provides intelligent, dynamic workload distribution to optimize resource use across the entire cluster. This is paired with intelligent resource allocation that maximizes GPU use and automates scaling across a broad ecosystem of accelerators from NVIDIA, AMD, and Intel as well as major OEM server partners.

Security and governance are woven into the entire lifecycle, providing a layered approach to safety and observability. Integrated tools track performance and detect data and concept drift, while AI guardrails make certain that models remain fair, transparent, and reliable. This rigorous governance, combined with the flexibility to deploy in air-gapped and disconnected environments, supports digital sovereignty and helps enterprises meet regulatory requirements. Red Hat AI Enterprise provides the control and consistency needed to turn AI initiatives into a strategic competitive advantage anywhere—from the public cloud to the edge.

Test and run AI models with Red Hat Enterprise Linux AI

Red Hat Enterprise Linux AI is a foundation model platform designed to run LLMs in individual server environments. Building on Red Hat AI Inference Server, it delivers an optimized, immutable, purpose-built appliance for inference. By using image mode for Red Hat Enterprise Linux, the platform packages the operating system and application together as a bootable image and hence places AI in a container-like deployment for bare metal. These images come with popular AI libraries—including PyTorch—along with essential network drivers and support for hardware-optimized accelerators, like NVIDIA, Intel, and AMD.

This integrated approach facilitates Day 1 operations by maximizing throughput and minimizing latency across the hybrid cloud. By using vLLM as its high-performance runtime and including an LLM compressor, organizations can significantly reduce compute costs while maintaining high model accuracy. This flexible, hardware-agnostic solution simplifies the path from PoC to production, providing a reliable environment for organizations to scale their AI enterprise initiatives in an environment of their choosing.

Extend the platform capabilities with Red Hat OpenShift AI

Red Hat OpenShift AI is an integrated AI platform for managing the lifecycles of predictive and gen AI models and delivering AI-based applications at scale across hybrid cloud environments. It unites data scientists and developers under IT oversight to develop, train, fine-tune, and manage models to help speed the journey from experimental AI applications to production. As a self-managed offering or fully managed cloud service, Red Hat OpenShift AI builds on the proven capabilities of Red Hat OpenShift to provide a trusted, consistent, and scalable environment for building, deploying, and monitoring AI/ML applications and models across on-site, public cloud, and edge environments. In partnership with our technology ecosystem, Red Hat OpenShift AI speeds innovation, increases operational consistency, and offers hybrid cloud flexibility—promoting transparency, freedom of choice, and responsible AI implementation.

Red Hat OpenShift AI empowers enterprise IT teams to align foundation models with relevant data through advanced RAG and fine-tuning patterns. It provides a consistent approach to MLOps and genAIOps, using automated AI pipelines and registries to accelerate time-to-value. By integrating vLLM, OpenShift AI delivers high-throughput inference performance that scales across distributed, multiple GPU systems, ensuring that gen AI applications remain responsive under heavy production loads.

To solve the complexity of scaling, the platform includes llm-d for intelligent workload distribution and hardware-optimized scheduling across large clusters. For production reliability, enhanced observability provides real-time detection of data drift and bias, alongside inference guardrails that evaluate inputs and outputs. These capabilities make models transparent and reliable whether deployed in the public cloud or within on-premise and air-gapped environments.

Explore a comprehensive partner ecosystem

Red Hat's partner ecosystem offers tools, services, and solutions designed for a wide range of AI use cases. Partners—including independent software vendors (ISVs), global systems integrators (GSIs), and cloud service providers—collaborate with Red Hat to integrate and certify advanced AI/ML technologies with Red Hat AI. These extensive relationships allow organizations to explore, select, and implement the most suitable technologies to find innovative AI solutions, ensuring flexibility and efficiency throughout your AI model and application lifecycles.

- **AI accelerator and hardware vendors:** Through partnerships with NVIDIA, AMD, and Intel, Red Hat AI provides an end-to-end platform that scales in the hybrid cloud. Our validation and certification programs verify that hardware is fully applied, while optimized workload management ensures efficient GPU use, maximizing performance for customers across OEM servers from partners such as Dell and Lenovo.
- **AI model providers:** Red Hat AI provides security-focused access and a controlled use of models, optimizing them through quantization allowing an efficient deployment and simplified connection of AI models with an organization's private data. Red Hat validates and optimizes models such as Meta (Llama), or IBM (Granite), or Mistral and makes them available in Hugging Face.

- **Cloud service providers:** Red Hat AI is optimized for major public clouds, including Amazon Web Services (AWS), Microsoft Azure, Google Cloud, and IBM Cloud. These partner cloud services provide the hybrid cloud scale and long-term support required for enterprise AI workloads, giving customers the ability to move from the initial Proof of Concept (PoC) to a fully deployed application.
- **Independent software vendors:** Red Hat AI empowers ISVs to accelerate go to market by simplifying the development of AI-enabled applications. AI-driven applications allow ISVs to focus on AI innovation across any model, hardware, or cloud, while optimizing inference for performance and accuracy with connections to an organization's data.
- **Global systems integrators:** Red Hat helps GSIs accelerate their customer's journey by offering high-value services. With Red Hat AI, GSIs can deploy global, scalable, and sovereign AI solutions such as AI factories—optimizing their return on investment (ROI) for large-scale projects by combining flexible inference with optimized data connections.

Success in action

Many customers are already experiencing the benefits and business outcomes of deploying Red Hat AI solutions. Four organizations share a brief overview about their AI journeys.

RTLZWEI

RTLZWEI, a mid-sized German broadcasting and digital media company, recognized that AI-driven innovation was essential to differentiate itself and deliver superior viewer experiences. To modernize its core multifunctional on-premise system and accelerate AI development, RTLZWEI adopted Red Hat AI. The unified, hybrid platform allows the company to develop, modernize, and build AI models at scale while maintaining strict data sovereignty and security standards. By centralizing DevOps and MLOps, RTLZWEI created an intelligent foundation for advanced use cases like automated video transcription and data forecasting for advertisers.[1]

Key outcomes:

- Accelerated AI innovation and efficiency by consolidating IT sprawl into one unified platform for application development and data science.
- Increased competitive advantage with high-performance inference, such as using vLLM to speed up video transcription and translation for international markets.
- Enhanced accuracy and cost-effectiveness by fine-tuning models on-premise with private data, resulting in a 33% reduction in word error rates for transcription.
- Streamlined operational agility through automated environment builds that allow developers to spin up full testing environments in a few clicks.

Read the case study to learn more about RTLZWEI's experience.

1 *Red Hat case study. "RTLZWEI hones competitive edge with Red Hat OpenShift AI." 22 Dec. 2025.*

Red Hat

Case study highlights

"Adding Red Hat OpenShift AI to our arsenal has given our data scientists greater autonomy and better standards."[2]

Ömer Uyar
CEO, Intertech

"As an invaluable AI-driven solution, Red Hat OpenShift AI provides a streamlined environment that enables our data scientists to build and deploy more robust and secure models."[2]

Okan Çetinkaya
CDO – CAO, DenizBank

"Platforms like Red Hat OpenShift AI allow us to keep data on-premises while accelerating model development and enabling business units to leverage AI capabilities in a flexible way."[3]

Serdar Gürbüz
General Manager,
Turkish Technology,
a Turkish Airlines subsidiary

DenizBank

Data scientists working at DenizBank, a prominent private bank in Türkiye and the 5th largest in the country, wanted to convert its existing workflow into a less manual process with a more standardized approach. The bank's IT subsidiary, Intertech, began a project to provide a model development environment with automated pipelines and standards to increase productivity and time to market. Intertech adopted Red Hat AI for its self-service capabilities and capacity to scale model serving and increase operational efficiency. Data scientists can now focus on building reliable and security-focused models that are even more effective.[2]

Key outcomes:

- Provided more than 120 data scientists from different lines of business greater autonomy and more consistent standards.
- Accelerated time to market while ensuring more reliable and security-focused models with automated environment builds and self-service capabilities.
- Optimized GPU use with slicing to maximize resource use, increase flexibility, and allow more workloads to run simultaneously without the need for additional GPU hardware.

Read the case study to learn more about DenizBank's experience.

Turkish Airlines

Turkish Airlines, the national flag carrier of Türkiye, launched a data-driven transformation program to gain a competitive edge in the global aviation industry. The airline's tech company, Turkish Technology, adopted Red Hat AI to empower business units as citizen data scientists and scale AI for more than 60 live models. This AI platform provides a standardized container orchestration layer that manages resource allocation for GPUs and automates workspace creation. By centralizing its AI initiatives, Turkish Airlines has moved from manual provisioning to a self-service model, fostering a culture of innovation that targets over US$100 million in financial benefits.[3]

Key outcomes:

- Accelerated model deployment by halving the time required to move AI projects into production through standardized templates and self-service tools.
- Increased operational efficiency by reducing workspace creation time from hours to minutes, allowing data scientists to begin experimentation almost instantly.
- Optimized resource use through automated GPU scaling and allocation, ensuring that intense computing tasks do not disrupt other workloads.
- Improved proactive decision-making with live models for real-time dynamic pricing, ground-time predictions, and payment fraud detection.

Read the case study to learn more about Turkish Airlines' experience.[3]

2 *Red Hat case study. "DenizBank empowers its data scientists." 16 Jan. 2025.*

3 *Red Hat press release. "Turkish Airlines Pioneers AI-led Innovation for Aviation with Red Hat OpenShift AI." 20 May 2025.*

Red Hat

Red Hat wanted to increase the efficiency and scalability of customer and technical support services for our growing customer base with AI solutions. The Experience Engineering team at Red Hat started working on a program—using Red Hat AI—to develop, test, and deploy 4 AI-powered solutions with the aim of simplifying IT support for its customers and support associates. The resulting tool helped improve the self-service process and increase efficiency, bringing about a faster customer response to support cases.[4]

Key outcomes:

- Delivered more than US$5 million in cost avoidance, with estimated US$1.5 million in 10 months.
- Increased availability of knowledge content and minimized repetitive tasks for IT support associates who handle 30,000 new cases each month.
- Using AI helped provide faster responses to customers, enhancing overall user experiences.

Read the case study to learn more about our IT support solution.

Learn more

Red Hat AI can assist organizations at any stage of their enterprise AI journey. Learn more about Red Hat AI solutions and how to get started.

4 *Red Hat case study. "Red Hat saves $5 million in IT support costs with AI augmentation." 17 Dec. 2024.*

About Red Hat

Red Hat helps customers standardize across environments, develop cloud-native applications, and integrate, automate, secure, and manage complex environments with award-winning support, training, and consulting services.

facebook.com/redhatinc
@RedHat
linkedin.com/company/red-hat

North America
1 888 REDHAT1
www.redhat.com

Europe, Middle East, and Africa
00800 7334 2835
europe@redhat.com

Asia Pacific
+65 6490 4200
apac@redhat.com

Latin America
+54 11 4329 7300
info-latam@redhat.com

redhat.com
#3402174_0226