# Red Hat Boosts Enterprise AI Across the Hybrid Cloud with Red Hat AI

Category: Business
written by News Mall | March 27, 2025
Red Hat, Inc., the worlds leading provider of open-source solutions, today announced the latest updates to Red Hat AI, its portfolio of products and services designed to help accelerate the development and deployment of AI solutions across the hybrid cloud. Red Hat AI provides an enterprise AI platform for model training and inference that delivers increased efficiency, a simplified experience and the flexibility to deploy anywhere across a hybrid cloud environment.

Even as businesses look for ways to reduce the costs of deploying large language models (LLMs) at scale to address a growing number of use cases, they are still faced with the challenge of integrating those models with their proprietary data that drives those use cases while also being able to access this data wherever it exists, whether in a data center, across public clouds or even at the edge.

Encompassing both Red Hat OpenShift AI and Red Hat Enterprise Linux AI (RHEL AI), Red Hat AI addresses these concerns by providing an enterprise AI platform that enables users to adopt more efficient and optimized models, tuned on business-specific data and that can then be deployed across the hybrid cloud for both training and inference on a wide-range of accelerated compute architectures.

**Red Hat OpenShift AI**

Red Hat OpenShift AI provides a complete AI platform for managing predictive and generative AI (gen AI) lifecycles

across the hybrid cloud, including machine learning operations (MLOps) and LLMOps capabilities. The platform provides the functionality to build predictive models and tune gen AI models, along with tools to simplify AI model management, from data science and model pipelines and model monitoring to governance and more.

Red Hat OpenShift AI 2.18, the latest release of the platform, adds new updates and capabilities to support Red Hat AI's aim of bringing better optimized and more efficient AI models to the hybrid cloud. Key features include:

- **Distributed serving**: Delivered through the vLLM inference server, distributed serving enables IT teams to split model serving across multiple graphical processing units (GPUs). This helps lessen the burden on any single server, speeds up training and fine-tuning and makes more efficient use of computing resources, all while helping distribute services across nodes for AI models.

- **An end-to-end model tuning experience**: Using InstructLab and Red Hat OpenShift AI data science pipelines, this new feature helps simplify the fine-tuning of LLMs, making them more scalable, efficient and auditable in large production environments while also delivering manageability through the Red Hat OpenShift AI dashboard.

- **AI Guardrails**: Red Hat OpenShift AI 2.18 helps improve LLM accuracy, performance, latency and transparency through a technology preview of AI Guardrails to monitor and better safeguard both user input interactions and model outputs. AI Guardrails offers additional detection points in helping IT teams identify and mitigate potentially hateful, abusive or profane speech, personally identifiable information, competitive information or other data limited by corporate policies.

- **Model evaluation**: Using the language model evaluation (lm-eval) component to provide important information on the model's overall quality, model evaluation enables data scientists to benchmark the performance of their LLMs across a variety of tasks, from logical and mathematical reasoning to adversarial natural language and more, ultimately helping to create more effective, responsive and tailored AI models.

**RHEL AI**

Part of the Red Hat AI portfolio, RHEL AI is a foundation model platform to more consistently develop, test and run LLMs to power enterprise applications. RHEL AI provides customers with Granite LLMs and InstructLab model alignment tools that are packaged as a bootable Red Hat Enterprise Linux server image and can be deployed across the hybrid cloud.

Launched in February 2025, RHEL 1.4 added several new enhancements including:

- **Granite 3.1 8B model support** for the latest addition to the open source-licensed Granite model family. The model adds multilingual support for inference and taxonomy/knowledge customization (developer preview) along with a 128k context window for improved summarization results and retrieval-augmented generation (RAG) tasks.

- **A new graphical user interface for skills and knowledge contributions**, available as a developer preview, to simplify data ingestion and chunking as well as how users add their own skills and contributions to an AI model.

- **Document Knowledge-bench (DK-bench)** for easier comparisons of AI models fine-tuned on relevant, private data with the performance of the same un-tuned base

models.

**Red Hat AI InstructLab on IBM Cloud**

Increasingly, enterprises are looking for AI solutions that prioritize accuracy and data security, while also keeping costs and complexity as low as possible. Red Hat AI InstructLab deployed as a service on IBM Cloud is designed to simplify, scale and help improve the security footprint for the training and deployment of AI models. By simplifying InstructLab model tuning, organizations can build more efficient models tailored to the organizations' unique needs while retaining control of their data.

**No-cost AI Foundations training**

AI is a transformative opportunity that is redefining how enterprises operate and compete. To support organizations in this dynamic landscape, Red Hat now offers AI Foundations online training courses at no cost. Red Hat is providing two AI learning certificates that are designed for experienced senior leaders and AI novices alike, helping educate users of all levels on how AI can help transform business operations, streamline decision-making and drive innovation. The AI Foundations training guides users on how to apply this knowledge when using Red Hat AI.

**Availability**

Red Hat OpenShift AI 2.18 and Red Hat Enterprise Linux AI 1.4 are now generally available. More information on additional features, improvements, bug fixes and how to upgrade to the latest version of Red Hat OpenShift AI can be found here and the latest version of RHEL AI can be found here.

Red Hat AI InstructLab on IBM Cloud will be available soon. AI Foundations training from Red Hat is available to customers now.

**Supporting Quotes**

**Joe Fernandes, vice president and general manager, AI Business Unit, Red Hat**, *"Red Hat knows that enterprises will need ways to manage the rising cost of their generative AI deployments, as they bring more use cases to production and run at scale. They also need to address the challenge of integrating AI models with private enterprise data and be able to deploy these models wherever their data may live. Red Hat AI helps enterprises address these challenges by enabling them to leverage more efficient, purpose-built models, trained on their data and enable flexible inference across on-premises, cloud and edge environments."*

**Regis Lesbarreres, advanced analytics and AI innovation manager, digital innovation, Airbus Helicopters**, "*At the outset of our AI journey, Airbus Helicopters was looking to integrate AI into our existing architecture, reduce shadow IT, unite our data scientists under a singular AI platform and optimize costs at scale. With Red Hat OpenShift AI, we've been able to accomplish all of these goals, which led to our first business use case for AI. Red Hat's vision for AI aligns with our business objectives and allows us to meet them while maintaining flexibility, accessibility and transparency."*

**Javier Olaizola Casin, Global Managing Partner, Hybrid Cloud and Data, IBM Consulting**, *"Businesses are increasingly applying AI to transform core business processes, and they need AI solutions that are flexible, cost-effective and tuned with trusted enterprise data to meet their unique needs. Red Hat AI brings the consistency, reliability and speed that organizations need to build and deploy AI models and applications across hybrid cloud scenarios. Combining IBM Consulting's domain, technology and industry expertise with Red Hat's AI technologies, we are helping our clients drive ROI from their technology investments."*

**Torsten Volks, Principal Analyst, Application Modernization,**

**ESG**, “Leading organizations harness AI-driven, data-centric decision making across teams and business units. Therefore, the ability to rapidly and cost effectively develop, deploy, integrate, scale and govern AI-based capabilities across the enterprise becomes a critical success factor. Establishing this capability requires an open and extensible AI foundation that ensures seamless integration with existing systems and processes, operational agility and continuous governance. Enabling staff and customers to benefit from AI capabilities faster and in a more comprehensive manner is crucial for continued business success.”

**Anand Swamy, Executive Vice President and Global Head of Ecosystems, HCLTech**, *“In order to realize the full potential of generative AI, organizations need to prioritize agile and flexible infrastructure. By combining the capabilities of Red Hat AI, encompassing RHEL AI and Red Hat OpenShift AI to deliver an end-to-end AI application platform, with HCLTech’s leading AI expertise and cognitive infrastructure services, a part of HCLTech AI Foundry solution, customers get a streamlined path to unlocking AI innovation, helping them overcome common challenges such as data security, scaling AI workloads and minimizing infrastructure costs.”*

**Additional Resources**

- Learn more about Red Hat AI
- Explore Red Hat’s AI portfolio
- Read more about Red Hat AI training

**About Red Hat, Inc.**

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 integrate new and existing IT applications, develop cloud-native applications, standardize on our industry-leading operating system, and automate, secure, and manage complex environments. Award-winning support, training, and consulting services make Red Hat a trusted adviser to the Fortune 500. As a strategic partner to cloud providers, system integrators, application vendors, customers, and open source communities, Red Hat can help organizations prepare for the digital future.

**Forward-Looking Statements**

Except for the historical information and discussions contained herein, statements contained in this press release may constitute forward-looking statements within the meaning of the Private Securities Litigation Reform Act of 1995. Forward-looking statements are based on the company's current assumptions regarding future business and financial performance. These statements involve a number of risks, uncertainties and other factors that could cause actual results to differ materially. Any forward-looking statement in this press release speaks only as of the date on which it is made. Except as required by law, the company assumes no obligation to update or revise any forward-looking statements.