

Red Hat AI Factory with NVIDIA

Scale AI from pilot to production with confidence

The challenge: AI that delivers real business value

Many businesses have moved beyond early AI development and are focused on operationalization and scale. In fact, nearly one-third of AI-enabled applications will rely on agentic AI by the end of 2026.¹ IT leaders driving this effort face:

Rising cost: AI workloads are unpredictable and inefficient, driving higher infrastructure spend, especially as inference scales.

Growing complexity: Disconnected tools, manual processes, and inconsistent environments slow deployment and make scaling AI difficult.

Limited control: Lack of centralized governance, security, and visibility creates risk across AI usage, performance, and compliance.

The solution: A repeatable, scalable, safeguarded factory process for AI delivery

Red Hat AI Factory with NVIDIA is a unified, co-engineered platform that combines Red Hat AI Enterprise and NVIDIA AI Enterprise to build, deploy, and scale production-ready AI across hybrid environments with confidence. With Red Hat AI Factory with NVIDIA, you can:



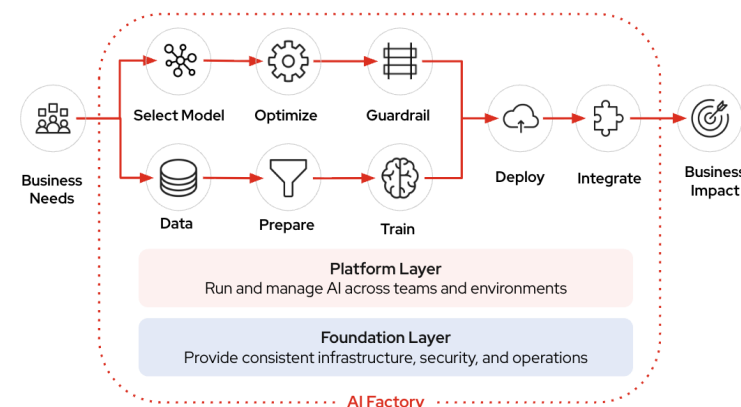
Integrate AI into your applications with full control over data, models, and deployment



Standardize how teams build and operate AI through a common platform to reduce complexity



Operate AI across environments on a consistent foundation without re-architecting, optimizing your costs



Better together

Brought to you by two of the most trusted names in the industry, **Red Hat AI Factory with NVIDIA** provides the power, flexibility, and security you need to confidently scale your AI.



Co-engineered for enterprise AI

- ▶ Jointly engineered platform aligned to evolving AI needs
- ▶ Curated, certified stack to scale AI workloads with confidence



Production-ready from Day 0

- ▶ Validated Blueprints and Quickstarts
- ▶ No need to design architectures from scratch



Consistent AI operations

- ▶ Single platform across data center, cloud, and edge
- ▶ Consistent security, governance, and control



Predictable, scalable economics

- ▶ Node-based pricing
- ▶ More predictable than GPU-based models
- ▶ Consistent way to buy and scale

Meaningful benefits

Red Hat AI Factory with NVIDIA is a unified, co-engineered platform that combines Red Hat AI Enterprise and NVIDIA AI Enterprise to build, deploy, and scale production-ready AI across hybrid environments with confidence. With Red Hat AI Factory with NVIDIA, you can:



Accelerate time to value Move from concept to production faster with validated workflows, blueprints, and optimized models



Increase operational efficiency Maximize utilization and reduce costs with optimized inference and scalable serving



Mitigate risk Secure, compliant, and trustworthy AI across hybrid environments

Services that deliver

- ▶ Accelerated path to production with expert guidance from Red Hat and trusted partners
- ▶ Reliable AI platform foundation
- ▶ Alignment to empowered teams on GenAIOps
- ▶ Proven business value through production use cases
- ▶ Strategic roadmap for scaling AI to production

Business-critical use cases



Model-as-a-service Centralized model hosting with APIs and multi-tenancy



Enterprise RAG Ground models in proprietary data with governance and orchestration



Enterprise research AI agents connect across enterprise data for deeper insight



Video search and summarization Analyze video data at scale using NVIDIA microservices

Make AI work for your business with Red Hat AI Factory with NVIDIA

Red Hat and NVIDIA have made it easier than ever for you to realize the power of AI. Purchased and delivered through trusted OEMs, solution providers, and distributors, and featuring new streamlined purchasing and lower costs, you can simplify procurement without sacrificing service.

Get started or learn more about Red Hat AI Factory with NVIDIA →



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.

f facebook.com/redhatinc
t @RedHat
in 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.

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