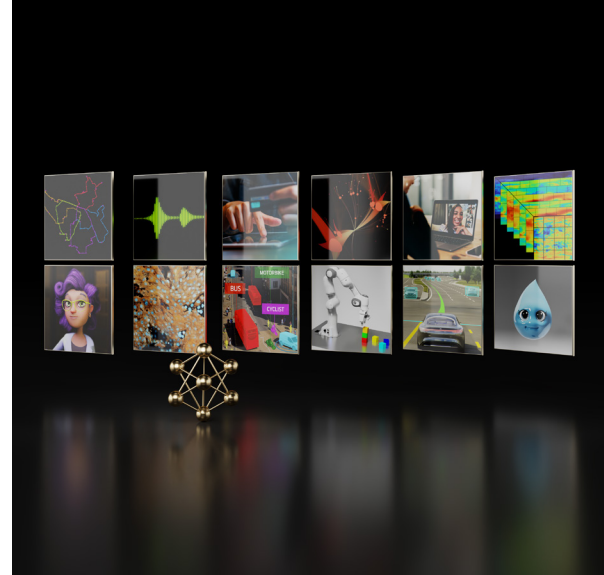




NVIDIA AI Enterprise

Build and accelerate production AI with a secure, reliable cloud-native software platform.



Unleashing AI for Every Enterprise

Artificial intelligence is transforming every industry, whether it's by improving customer relationships in financial services, streamlining manufacturer supply chains, or helping doctors deliver better outcomes for patients. IDC projects that, by 2024, 60 percent of the Global 2000¹ will expand the use of AI and machine learning across all business-critical horizontal functions, such as marketing, legal, HR, procurement, and supply chain logistics.

While most organizations understand the potential of AI to transform their business and keep them competitive, the technology is fairly unfamiliar to many IT and line-of-business stakeholders.

Discover the Platform for Building Production AI

NVIDIA AI Enterprise is an end-to-end, secure, cloud-native suite of AI software that enables organizations to solve new challenges while increasing operational efficiency. It accelerates the data science pipeline and streamlines the development and deployment of predictive AI models to automate essential processes and gain rapid insights from data. With an extensive library of full-stack software, including AI solution workflows, frameworks, pretrained models, and infrastructure optimization, the possibilities are endless.

Available in the cloud, in the data center, and at the edge, NVIDIA AI Enterprise lets organizations develop once and run anywhere. Global enterprise support and regular security reviews ensure business continuity and AI projects stay on track.

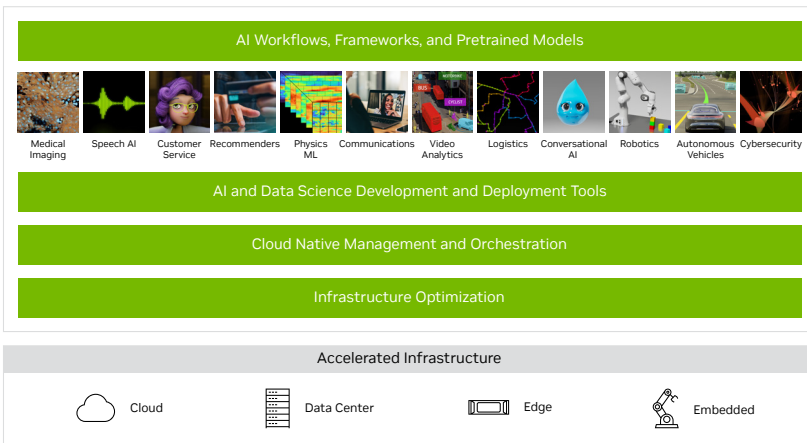
NVIDIA AI Workflows

AI workflows are cloud-native, prepackaged reference examples that illustrate how NVIDIA AI frameworks can be leveraged to build AI solutions, including intelligent virtual assistants, cybersecurity digital fingerprinting for anomaly detection, product recommendations, and more. With pretrained models, training and inference pipelines, Jupyter Notebooks, and Helm charts, NVIDIA's AI workflows accelerate the path to delivering AI outcomes. AI workflows can run as microservices and can be deployed on Kubernetes alone or with other microservices to create production-ready applications. Reduce time to deployment, lower costs, improve accuracy and performance, and have more confidence in AI outcomes with NVIDIA AI Enterprise's packaged workflows, supported by NVIDIA AI experts.

AI Adoption Challenges

- > **Risk:** Pulling together an end-to-end AI solution from disparate products—and integrating them with existing infrastructure—is difficult.
- > **Performance:** Fast time to deployment and high performance are critical for AI, machine learning, and data analytics workloads.
- > **Scaling:** Going from proof of concept to enterprise deployment requires effective scaling through efficient use of resources to ensure manageability, availability, and infrastructure cost management.

NVIDIA AI Enterprise
End-to-end Platform for Production AI



Benefits of NVIDIA AI Enterprise

- > **Optimized to simplify AI** development and deployment with included AI frameworks and containers, enabling insights to be gathered faster and business value to be delivered sooner.
- > **Certified to deploy anywhere**, including in the cloud, on popular data center platforms from VMware and Red Hat, and on NVIDIA-Certified Systems™.
- > **Supported by NVIDIA AI experts** so projects stay on track, with included priority notifications, long-term support, and customized support-upgrade options.

Optimize Every Step of the AI Workflow

NVIDIA AI Enterprise's tools and frameworks streamline and support every stage of the AI journey, from data prep and model training through inference, and deployment at scale:

- > **Speed data processing time** by up to 5X, while reducing operational costs by up to 5X, over CPU-only platforms with the **NVIDIA RAPIDS™ Accelerator for Apache Spark**.
- > **Train at scale** with the **NVIDIA TAO Toolkit**. Create custom, production-ready AI models in hours, rather than months, by fine-tuning NVIDIA pretrained models—without AI expertise or large training datasets.
- > **Optimize for inference** with **NVIDIA® TensorRT™**-based applications that perform up to 40X faster than CPU-only platforms. With TensorRT, fine-tune neural network models trained in all major frameworks.
- > **Deploy at scale** with **NVIDIA Triton™ Inference Server**, which simplifies and optimizes the deployment of AI models at scale and in production for both neural networks and tree-based models on GPUs.

Explore Licensing Options

Purchased as either perpetual licenses with three- or five-year support services included or as a subscription. Learn more about **product pricing and licensing**.

Included with NVIDIA H100 Tensor Core GPUs for mainstream servers. Activation is required.

Included in NVIDIA DGX™ software.

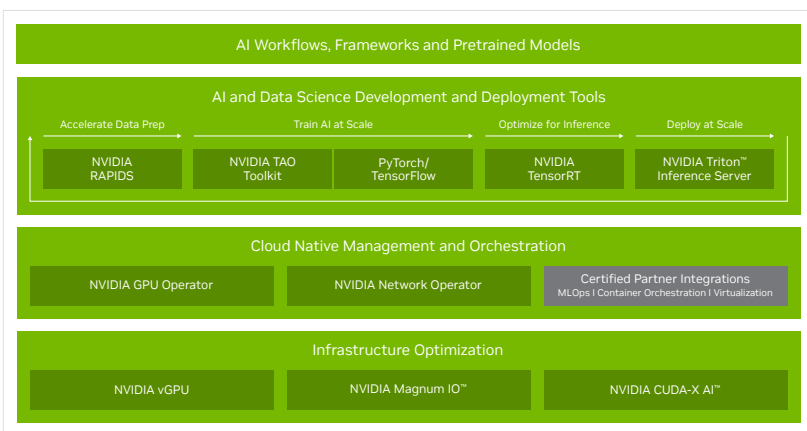
Get Hands-On Experience

- > **NVIDIA LaunchPad** offers free, hands-on labs so you can experience how to create enterprise AI, including intelligent chatbots, recommendation engines, image classification, and more.

Sign up for an Evaluation

- > A **ninety-day evaluation** license lets you start your proof of concepts.

NVIDIA AI Enterprise Software Suite
Streamlined AI Development and Deployment



Get Enterprise-Grade Security and Support

With NVIDIA AI Enterprise, organizations get the transparency of open source with the assurance of NVIDIA support when they move from development to production:

- > Full enterprise support for NVIDIA AI application frameworks, NVIDIA pretrained models, and other NVIDIA AI software designated as eligible for support and available on NVIDIA NGC™, a portal of enterprise services, software, and management tools.
- > Support provided by global NVIDIA AI experts and NVIDIA engineering teams through defined escalation paths
- > Guaranteed service-level agreements
- > Priority notifications related to the latest security fixes and maintenance releases
- > Long-term support for up to three years for designated software branches
- > Customized support-upgrade options, including a designated technical account manager and business-critical support for 24/7 live agent access

Develop Once and Deploy Anywhere

NVIDIA AI Enterprise offers a unified platform to run NVIDIA AI-enabled solutions across certified multi-cloud, hybrid-cloud, and edge environments, making AI deployment scalable and flexible:

- > **Multi-cloud:** Deploy on GPU-accelerated public cloud instances, including AWS, Azure, Google Cloud, and Oracle Cloud Infrastructure. NVIDIA AI Enterprise is also available on major cloud marketplaces.
- > **Hybrid cloud:** Run on enterprise platforms, including VMware Cloud Foundation, Red Hat Enterprise Linux, and HPE GreenLake.
- > **Container orchestration:** Use with mainstream container platforms, including VMware Tanzu, Red Hat OpenShift, HPE Ezmeral, Google Kubernetes Engine (GKE), Amazon Elastic Kubernetes Service (EKS), and upstream Kubernetes.
- > **NVIDIA DGX systems:** NVIDIA AI Enterprise software is included with the DGX software to streamline AI development and deployment.
- > **NVIDIA Certified-Systems:** NVIDIA AI Enterprise is certified to run on over 400 NVIDIA-Certified Systems available from a wide range of equipment manufacturers. A five-year license of NVIDIA AI Enterprise is included with the NVIDIA H100 PCIe Tensor Core GPU for mainstream servers.

Tap Into NVIDIA's Ecosystem of Partners

A broad ecosystem of certified partner integrations reduces deployment risk:

- > MLOps solution providers for collaboration and productivity
- > Global solution integrators for customized state-of-the-art AI solutions
- > Service delivery partners for applied AI services

Ready to Get Started?

To learn more about the NVIDIA AI solutions for enterprise, visit:

nvidia.com/ai-enterprise-suite

nvidia.com/try-ai

nvidia.com/certified-systems

¹ IDC FutureScape: Worldwide Artificial Intelligence and Automation 2022 Predictions

Partner
Logo

