AI-Ready Enterprise Platform
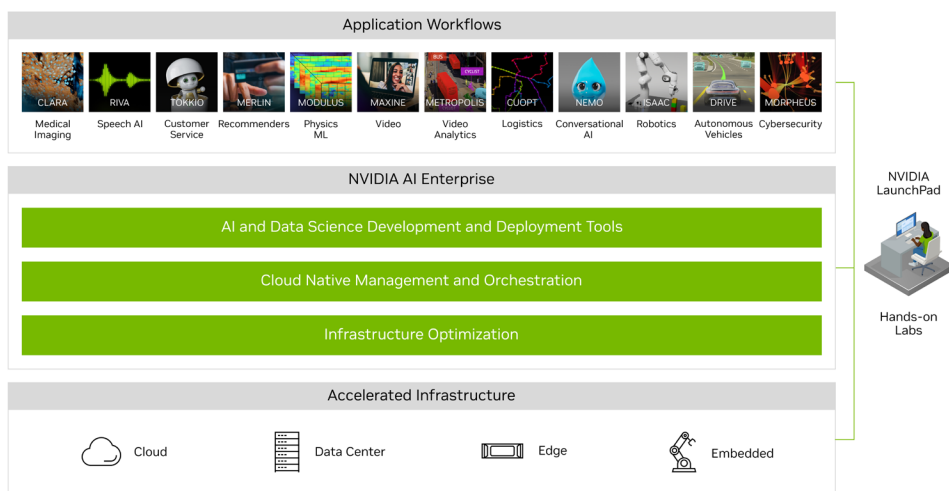
Unleashing AI for every enterprise.

Artificial intelligence (AI) 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% of the G2000 will expand the use of AI and machine learning (ML) 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 still fairly unfamiliar to many IT and line-of-business stakeholders.

Discover the AI-Ready Platform from NVIDIA.

NVIDIA AI Enterprise is an end-to-end, cloud-native suite of AI and data analytics software, optimized so every organization can excel at AI.

As the operating system of the NVIDIA AI platform, NVIDIA AI Enterprise is essential for production and support of applications built with the extensive NVIDIA library of frameworks and pretrained models, such as NVIDIA Riva® for speech AI, NVIDIA Merlin™ for recommendation engines, NVIDIA Clara™ for medical imaging, and more. NVIDIA AI Enterprise is certified to deploy on broadly adopted enterprise platforms, including multi-cloud instances on AWS, Microsoft Azure, and Google Cloud as well as NVIDIA-Certified Systems™ from leading server vendors.

KEY CHALLENGES

> Key Challenges for Enterprise AI Workloads

> **Risk** - Pulling together an end-to-end AI solution from disparate products, and integrating them with existing infrastructure, is difficult.

> **Performance** - Fast deploy and high performance are critical for AI, ML, 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.
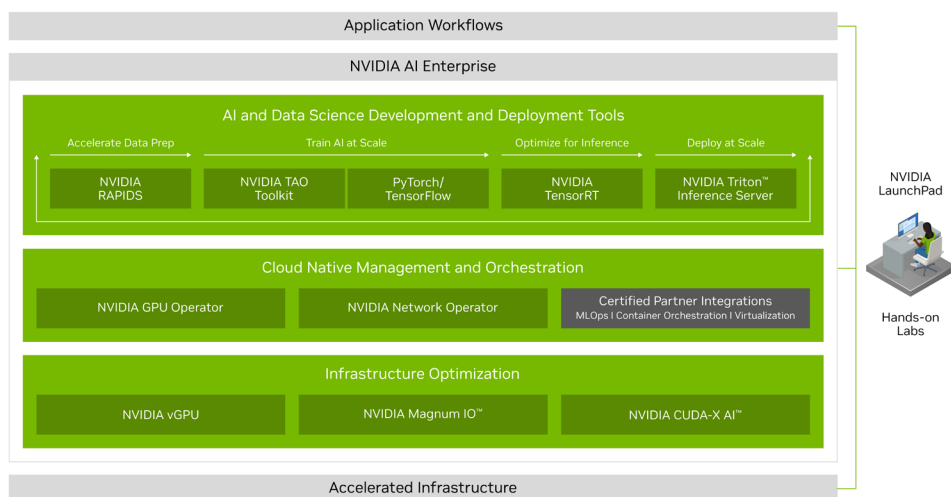
BENEFITS FOR ENTERPRISES

> **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.

NVIDIA's latest data center flagship products, the NVIDIA DGX and mainstream H100 Tensor Core GPUs, include NVIDIA AI Enterprise to ensure organizations have access to state-of-the-art AI software.

Optimize every step of the AI workflow, including data prep, model training, inference, and deployment at scale with NVIDIA AI tools and frameworks:

> Accelerate data prep with **NVIDIA RAPIDS™**. Enterprises can achieve up to 70X faster performance and up to 20X more cost-effectiveness compared to similar CPU-only configurations.

> 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.

A broad ecosystem of certified partner integrations reduces deployment risk:

> MLOps solution providers for collaboration and productivity

> VMware vSphere, VMware Cloud Foundation, and VMware Cloud Director to scale in virtualized environments

> Red Hat OpenShift certification

Get support for production AI. Organizations get the transparency of open source with the assurance of support from NVIDIA when they move from development to production:

> Full enterprise support for conversational AI, recommendation engines, medical imaging, and more AI workflows

> Access to NVIDIA AI experts and engineering teams

> Priority notifications related to the latest security fixes and maintenance releases

> **Enterprise Training Services** included so developers, data scientists, and IT professionals get the most out of NVIDIA AI Enterprise

> 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

Rely on NVIDIA-Certified Systems.

> The NVIDIA AI Enterprise software suite is certified to run on NVIDIA-Certified Systems, which include the following:

> NVIDIA Ampere and Hopper architecture-based GPUs

> NVIDIA ConnectX® SmartNICs and NVIDIA BlueField® data processing units (DPUs)

> NVIDIA converged accelerators, which combine the powerful performance of the NVIDIA GPU architecture with the enhanced security and latency-reduction capabilities of the NVIDIA BlueField-2 DPU.

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

DELTA
C O M P U T E R

nVIDIA.