NVIDIA AI Enterprise

Features

- Comprehensive AI development suite, covering data preparation, training, AI inference library and services
- Lifecycle management & security update, including 2 production branches per year with a 9-month lifetime and monthly CVE patches & bug fixes.
- Expert support and services, including Advantech Tier 1 services for AI development infrastructure: OS, drivers & AI SDKs, and direct access to Nvidia AI experts

Enterprise-Grade AI Software Platform From NVIDIA

NVIDIA AI Enterprise is an end-to-end, cloud-native software platform that accelerates data science pipelines and streamlines development and deployment of production-grade co-pilots and other generative AI applications. Easy-to-use microservices provide optimized model performance with enterprise-grade security, support, and stability to ensure a smooth transition from prototype to production for enterprises that run their businesses on AI.

Enterprise-Grade Security, Stability, Manageability, and Support

As AI rapidly evolves and expands, the software stack supporting it becomes more complex as it becomes more relied upon. NVIDIA AI Enterprise addresses this challenge providing a platform specifically designed for running mission-critical AI applications in businesses. It offers regular releases of security patches to address critical vulnerabilities and exposures (CVEs), maintains a production branch for API stability, provides comprehensive end-to-end management software, and offers enterprise-level support with service-level agreements (SLAs).

NVIDIA AI Software Supported

All NVIDIA AI application frameworks, pretrained models, and other NVIDIA AI software accessible through NGC are covered by an NVIDIA AI Enterprise license. This includes over 100 AI frameworks and pretrained models such as NeMo, Maxine, cuOpt, and more (with additions ongoing). To identify supported software, look for the "NVIDIA AI Enterprise Supported" label on NGC.
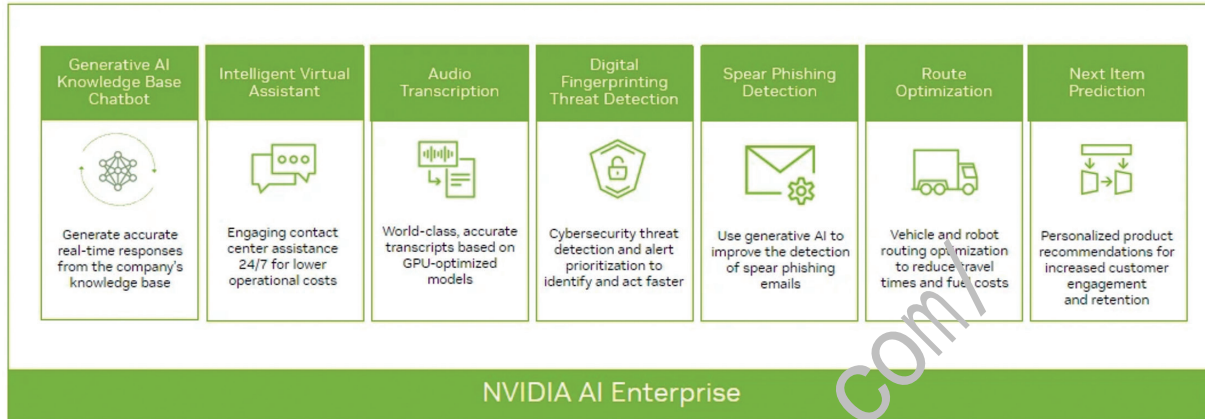
NVIDIA AI Workflows

NVIDIA AI Enterprise includes new AI solution workflows for building AI applications, including intelligent virtual assistants, audio transcription, and cybersecurity digital fingerprinting to detect anomalies. These packaged AI workflow examples include NVIDIA AI frameworks and pretrained models, as well as resources such as Helm Charts, Jupyter Notebooks, and documentation to help customers more easily get started in building AI-based solutions.

Production-Ready Pretrained Models

NVIDIA AI Enterprise provides access to pretrained models tailored for healthcare and vision AI tasks, including people detection, vehicle detection, federated learning models, image registration, and more. These pretrained models are not encrypted, allowing developers to examine each model's weights and biases. This transparency aids in model explainability and understanding potential biases. Additionally, unencrypted models are easier to debug and integrate into custom AI applications.

NVIDIA NIM

NIM is a collection of microservices designed to accelerate deployment of generative AI across your enterprise. Leveraging industry standard APIs, developers can quickly build enterprise-grade AI applications with just a few lines of code. Built on robust foundations including inference engines like Triton Inference Server, TensorRT, TensorRTLLM, and PyTorch, NIM is engineered to facilitate seamless AI inferencing at scale.

NVIDIA AI Enterprise Support Services

NVIDIA AI Enterprise customers receive live expert guidance with response times covered in the service-level-agreement (SLA). Customers are also provided with access to a case management portal, software updates, maintenance releases, and security notifications. This comprehensive support ensures that businesses can rely on NVIDIA AI experts to keep their AI suite running smoothly and help them reach their AI objectives.

Support Options	Services	Supported Channels	Service Level Agreement
Business Standard (required for perpetual)	Issue resolution, bug fixes, software updates, maintenance	Phone, portal, email	9 x 5 business days (4-hour initial response)
Business Critical (optional upgrades/ add-ons)	Same as Business Standard Support	Same as Business Standard Support	24 x 7 (1-hour initial response for S1 cases)

System requirements

Supported Hardware

NVIDIA AI Enterprise is supported on the following NVIDIA GPUs with the compatible Advantech systems that are listed on the NVIDIA-certified systems page.

NVIDIA A800 PCIe 80GB	NVIDIA A30 liquid cooled	NVIDIA H100 SXM5 94GB	NVIDIA L40S
NVIDIA A800 PCIe 80GB liquid cooled	NVIDIA A30X	NVIDIA H100 SXM5 80GB3	NVIDIA RTX A6000
NVIDIA A800 HGX 80GB	NVIDIA A30	NVIDIA H100 SXM5 64GB	NVIDIA RTX A5500
NVIDIA A800 40GB PCIe active cooled	NVIDIA A10	NVIDIA H800 PCIe 94GB (H800 NVL)	NVIDIA RTX A5000
NVIDIA A100X	NVIDIA A16	NVIDIA H800 PCIe 94GB	NVIDIA RTX 6000 passive
NVIDIA A100 PCIe 40GB	NVIDIA A2	NVIDIA H800 PCIe 80GB	NVIDIA RTX 8000 passive
NVIDIA A100 HGX 40GB	NVIDIA AX8001	NVIDIA H800 SXM5 80GB3	NVIDIA RTX 6000 Ada
NVIDIA A100 PCIe 80GB	NVIDIA GH200 96GB (CG1) Grace Hopper™ Superchip2	NVIDIA L2	NVIDIA RTX 5880 Ada
NVIDIA A100 PCIe 80GB liquid cooled	NVIDIA GH200 Grace Hopper 144GB (CG1) Superchip2	NVIDIA L4	NVIDIA RTX 5000 Ada
NVIDIA A100 HGX 80GB	NVIDIA H100 PCIe 94GB (H100 NVL)	NVIDIA L20	NVIDIA T4
NVIDIA A40	NVIDIA H100 PCIe 80GB	NVIDIA L40	NVIDIA V100

Recommended Operating System

- Ubuntu LTS releases (20.04 or above)

Ordering information

- **310TNVAIS101A1**   Subscription, 1-year term, Business Standard Support, per GPU
- **310TNVAIS301A1**   Subscription, 3-year term, Business Standard Support, per GPU
- **310TNVAIS501A1**   Subscription, 5-year term, Business Standard Support, per GPU
- **310TNVAIS001A1**   Perpetual License + 5-year Business Standard Support, per GPU