License for the AI Autobahn: NVIDIA AI Enterprise 3.0 Introduces New Tools to Speed Success

Software suite to offer AI workflows for contact center intelligent virtual assistants, audio transcription, digital fingerprinting for cybersecurity, as well as support for 50+ NVIDIA AI frameworks and pretrained models.

Author: Justin Boitano

From rapidly fluctuating demand to staffing shortages and supply chain complexity, enterprises have navigated numerous challenges the past few years. Many companies seeking strong starts to 2023 are planning to use AI and accelerated computing to drive growth while saving costs.

To support these early adopters — as well as those just beginning their AI journey — NVIDIA has announced a new version of its NVIDIA AI Enterprise software suite to support businesses worldwide across a wide range of domain and industry-specific workloads.

NVIDIA AI Enterprise 3.0 will introduce workflows for contact center intelligent virtual assistants, audio transcription and digital fingerprinting for cybersecurity — some of the most common applications for enterprises adopting AI to better serve customers.

Expected to be available later this month, NVIDIA AI Enterprise 3.0 also expands support for more than 50 NVIDIA AI software frameworks and pretrained models available on the NVIDIA NGC software catalog, supercharging and simplifying AI deployments for organizations globally.

The new software release arrives as Deutsche Bank today announced its plans to partner with NVIDIA to accelerate the use of AI in financial services as part of its strategy for developing AI-powered speech, vision and fraud detection applications within the industry.

"AI and machine learning will redefine banking, and we're already working closely with NVIDIA to lead the industry in leveraging these technologies to improve customer service and mitigate risk," said Gil Perez, chief innovation officer and head of Cloud & Innovation Network at Deutsche Bank. "Accelerated computing enables traders to manage risk and run more scenarios faster while also improving energy efficiency, and NVIDIA AI Enterprise provides the flexibility to support AI development across our hybrid infrastructure."

This latest version of our secure, cloud-native suite of AI software enables organizations to solve business challenges while increasing operational efficiency. It accelerates the data science pipeline and streamlines the development and deployment of AI models to automate essential processes and gain rapid insights from data.

The new AI workflows for contact center intelligent virtual assistants, audio transcription and cybersecurity digital fingerprinting in NVIDIA AI Enterprise 3.0 leverage NVIDIA expertise to reduce development time and costs to speed time to deployment.

The workflows run as cloud-native microservices using NVIDIA AI frameworks and pretrained models, as well as Helm charts, Jupyter notebooks and more. Enterprises can deploy the microservices as standalone Kubernetes containers or combine them with other services to create production-ready applications with greater accuracy and performance.

The NVIDIA contact center intelligent virtual assistant AI workflow enables enterprises to respond to customers around the clock to reduce wait times and free up time for human contact center agents to support more complex inquiries — all while reducing costs. Using the workflow, enterprises can develop agents that deliver personalized and precise responses in natural-sounding voices. By

leveraging AI, the agents can better understand customers even on calls with poor audio quality.

With the NVIDIA audio transcription AI workflow , enterprises can rapidly create accurate transcripts in English, Spanish, Mandarin, Hindi, Russian, Korean, German, French and Portuguese using NVIDIA automatic speech recognition technology, with Japanese, Arabic and Italian expected to be added soon. The transcription workflow leverages fully customizable GPU-optimized models to enable better understanding, contextual insights and sentiment analysis with real-time accuracy. Enterprises can use the completed transcripts to improve product development and speed training of contact center agents.

Using unsupervised learning, the NVIDIA digital fingerprinting AI workflow employs threat detection to achieve comprehensive data visibility. It improves security by helping enterprises uniquely fingerprint every user, service, account and machine across the network to detect anomalous behavior. Once deployed, the workflow provides intelligent alerts and actionable information to reduce detection time from weeks to minutes to help security analysts quickly identify and act on threats.

NVIDIA AI Enterprise 3.0 also features unencrypted pretrained models and source code from the latest release of NVIDIA TAO Toolkit , a low-code AI development solution for creating highly accurate, customized, production-ready AI models for speech and computer vision AI applications.

The unencrypted models are exclusively available with NVIDIA AI Enterprise and support a variety of imaging and vision AI tasks for healthcare, smart cities and retail, such as pathology tumor detection, people detection, vehicle detection, pose estimation and action recognition.

Using the pretrained models without encryption enables developers to view the weights and biases of the model, which can help in model explainability and understanding model bias. In addition, unencrypted models are easier to debug and easier to integrate into custom AI applications.

NVIDIA AI Enterprise 3.0 also introduces support for a broad range of NVIDIA AI frameworks and infrastructure options:

NVIDIA Clara Parabricks and MONAI improve healthcare: New support for NVIDIA Clara Parabricks enables faster, more accurate genomic analysis for sequencing centers, clinical labs, genomics researchers and genomics instrument manufacturers. NVIDIA AI Enterprise also supports MONAI , a domain-specific medical imaging AI framework that provides pretrained models and a collaborative, scalable workflow for data labeling and training robust AI models.

NVIDIA AI frameworks to boost customer service, safety, sales and more: The 50+ frameworks and pretrained models now supported in NVIDIA AI Enterprise 3.0 include NVIDIA Riva , a GPU-accelerated speech AI software development kit for building and deploying fully customizable, real-time AI pipelines that deliver world-class accuracy in all leading clouds, on premises, at the edge and on embedded devices. NVIDIA Morpheus enables cybersecurity developers to create optimized AI pipelines for filtering, processing and classifying large volumes of real-time data. SDKs in the NVIDIA Metropolis intelligent video analytics platform, such as TAO Toolkit and NVIDIA DeepStream for vision AI, are supported, as is the NVIDIA Merlin open-source framework for building high-performing recommender systems at scale.

Expanded certification for the cloud: With NVIDIA AI Enterprise 3.0, organizations with a hybrid cloud strategy now have the flexibility to run the software on GPU-accelerated instances from Oracle Cloud Infrastructure . Customers who purchase a license through one of NVIDIA's channel partners can deploy in OCI with full certification and support from NVIDIA on designated OCI instances. This is in addition to existing NVIDIA AI Enterprise certification for accelerated instances on Amazon Web Services, Google Cloud and Microsoft Azure.

Hewlett Packard Enterprise and NVIDIA extend AI support for hybrid data centers : HPE and NVIDIA will deliver a joint offering that provides support for the NVIDIA AI Enterprise 3.0 on HPE GreenLake and HPE Ezmeral. The solution allows customers to speed up AI application development, securely, by easily procuring and deploying NVIDIA AI Enterprise on a managed HPE GreenLake instance.

Broadened storage and virtualization support : NVIDIA AI Enterprise 3.0 now supports NVIDIA Magnum IO GPUDirect Storage , which provides a direct data path between local or remote storage and GPU memory to further speed AI workloads. It also delivers expanded virtualization options, including Red Hat Enterprise Linux with KVM and VMware vSphere 8 .

NVIDIA AI Enterprise is available now. Customers can contact NVIDIA partners worldwide for pricing. NVIDIA AI Enterprise 3.0 is expected to be available for customers with current and new subscriptions later this month. A license for NVIDIA AI Enterprise is also included with servers from NVIDIA partners that feature NVIDIA H100 PCIe GPUs, including systems from Dell Technologies, Hewlett Packard Enterprise, Lenovo and Supermicro.

Enterprises can grow their AI expertise by trying NVIDIA AI workflows and frameworks supported in NVIDIA AI Enterprise on NVIDIA LaunchPad at no charge.

Original URL: https://blogs.nvidia.com/blog/2022/12/07/ai-enterprise-software-3/