NVIDIA Healthcare Launches Generative AI Microservices to Advance Drug Discovery, MedTech and Digital Health

New Catalog of NVIDIA NIM and GPU-Accelerated Microservices for Biology, Chemistry, Imaging and Healthcare Data Runs in Every NVIDIA DGX Cloud

GTC—NVIDIA today launched more than two dozen new microservices that allow healthcare enterprises worldwide to take advantage of the latest advances in generative AI from anywhere and on any cloud.

The new suite of NVIDIA healthcare microservices includes optimized NVIDIA NIM™ AI models and workflows with industry-standard APIs, or application programming interfaces, to serve as building blocks for creating and deploying cloud-native applications. They offer advanced imaging, natural language and speech recognition, and digital biology generation, prediction and simulation.

Additionally, NVIDIA accelerated software development kits and tools, including Parabricks®, MONAI, NeMo™, Riva and Metropolis, can now be accessed as NVIDIA CUDA-X™ microservices to accelerate healthcare workflows for drug discovery, medical imaging and genomics analysis.

The microservices, 25 of which launched today, can accelerate transformation for healthcare companies as generative AI introduces numerous opportunities for pharmaceutical companies, doctors and hospitals. These include screening for trillions of drug compounds to advance medicine, gathering better patient data to aid early disease detection and implementing smarter digital assistants.

Researchers, developers and practitioners can use the microservices to easily integrate AI into new and existing applications and run them anywhere — from the cloud to on premises — equipping them with copilot capabilities to enhance their life-saving work.

“For the first time in history, we can represent the world of biology and chemistry in a computer, making computer-aided drug discovery possible,” said Kimberly Powell, vice president of healthcare at NVIDIA. “By helping healthcare companies easily build and manage AI solutions, we’re enabling them to harness the full power and potential of generative AI.”

NVIDIA NIM Healthcare Microservices for Inferencing
The new suite of healthcare microservices includes NVIDIA NIM, which provides optimized inference for a growing collection of models across imaging, medtech, drug discovery and digital health. These can be used for generative biology and chemistry, and molecular prediction. NIM microservices are available through the NVIDIA AI Enterprise 5.0 software platform.

The microservices also include a collection of models for drug discovery, including MolMIM for generative chemistry, ESMFold for protein structure prediction and DiffDock to help researchers understand how drug molecules will interact with targets. The VISTA 3D microservice accelerates the creation of 3D segmentation models. The Universal DeepVariant microservice delivers over 50x speed improvement for variant calling in genomic analysis workflows compared to the vanilla DeepVariant implementation running on CPU.

Cadence, a leading computational software company, is integrating NVIDIA BioNeMo™ microservices for AI-guided molecular discovery and lead optimization into its Orion® molecular design platform, which is used for accelerating drug discovery.

Orion allows researchers at pharmaceutical companies to generate, search and model data libraries with hundreds of billions of compounds. BioNeMo microservices, such as the MolMIM generative chemistry model and the AlphaFold-2 model for protein folding, substantially augment Orion’s design capabilities.

“Our pharmaceutical and biotechnology customers require access to accelerated resources for molecular simulation,” said Anthony Nicholls, corporate vice president at Cadence. “By leveraging BioNeMo microservices, researchers can generate molecules that are optimized according to scientists’ specific needs.”

Nearly 50 application providers are using the healthcare microservices, as are biotech and pharma companies and platforms, including Amgen, Astellas, DNA Nexus, Iambic Therapeutics, Recursion and Terray, and medical imaging software makers such as V7.

"Generative AI is transforming drug discovery by allowing us to build sophisticated models and seamlessly integrate AI into the antibody design process,” said David M. Reese, executive vice president and chief technology officer at Amgen. “Our

team is harnessing this technology to create the next generation of medicines that will bring the most value to patients."

Improving Patient and Clinician Interactions
Generative AI is changing the future of patient care. Hippocratic AI is developing task-specific Generative AI Healthcare Agents, powered by the company's safety-focused LLM for healthcare, connected to NVIDIA Avatar Cloud Engine microservices and will utilize NVIDIA NIM for low-latency inferencing and speech recognition.

These agents talk to patients on the phone to schedule appointments, conduct pre-operative outreach, perform post-discharge follow-ups and more.

"With generative AI, we have the opportunity to address some of the most pressing needs of the healthcare industry. We can help mitigate widespread staffing shortages and increase access to high-quality care — all while improving outcomes for patients," said Munjal Shah, cofounder and CEO of Hippocratic AI. "NVIDIA's technology stack is critical to achieving the conversational speed and fluidity necessary for patients to naturally build an emotional connection with Hippocratic's Generative AI Healthcare Agents."

Abridge is building an AI-powered clinical conversation platform that generates notes drafts, saving clinicians up to three hours a day. Going from raw audio in noisy environments to draft documentation requires many AI technologies to work together seamlessly. Language identification, transcription, alignment and diarization must all take place within seconds and conversations must be structured according to the sorts of medical information contained in each utterance, and powerful language models must be applied to transform the relevant evidence into summaries. The system turns clinical conversations into high-quality, after-visit documentation in real time.

Flywheel creates models that can be transformed into microservices. The company's centralized, cloud-based platform powers biopharma companies, life science organizations, healthcare providers and academic medical centers, helping them identify, curate and train medical imaging data to accelerate time to insight.

"In this rapidly evolving landscape of healthcare technology, the integration of NVIDIA's generative AI microservices with Flywheel's platform represents a transformative leap forward," said Trent Norris, chief product officer at Flywheel. "By leveraging these advanced tools, we are not only enhancing our capabilities in medical imaging and data management but also driving unprecedented acceleration in medical research and patient care outcomes. Flywheel's AI Factory powered by NVIDIA's cutting-edge AI solutions meets healthcare customers where they are, pushing the boundaries of what's possible in the realm of digital health and biopharma."

Availability
Developers can experiment with NVIDIA AI microservices at ai.nvidia.com and deploy production-grade NIM microservices through NVIDIA AI Enterprise 5.0 running on NVIDIA-Certified Systems™ from providers including Dell Technologies, Hewlett Packard Enterprise, Lenovo and Supermicro, leading public cloud platforms including Amazon Web Services (AWS), Google Cloud, Microsoft Azure and Oracle Cloud Infrastructure, and on NVIDIA DGX™ Cloud.

For more information, visit NVIDIA's booth at GTC, running March 18-21 at the San Jose Convention Center and online, and watch the replay of NVIDIA founder and CEO Jensen Huang's keynote.

About NVIDIA
Since its founding in 1993, NVIDIA (NASDAQ: NVDA) has been a pioneer in accelerated computing. The company's invention of the GPU in 1999 sparked the growth of the PC gaming market, redefined computer graphics, ignited the era of modern AI and is fueling industrial digitalization across markets. NVIDIA is now a full-stack computing infrastructure company with data-center-scale offerings that are reshaping industry. More information at https://nvidianews.nvidia.com/.

Certain statements in this press release including, but not limited to, statements as to: the benefits, impact, performance, features, and availability of NVIDIA's products and technologies, including NVIDIA accelerated software development kits and tools including Parabricks, MONAI, NeMo, Riva and Metropolis, NVIDIA AI Enterprise 5.0 software platform, NVIDIA's microservices, such as NIM, CUDA-X, BioNeMo, Avatar Cloud Engine, VISTA 3D, Universal DeepVariant, NVIDIA-Certified Systems; and NVIDIA DGX Cloud; our ability to represent the world of biology and chemistry in a computer, making computer-aided drug discovery possible; NVIDIA enabling healthcare companies to harness the full power and potential of generative AI by helping them easily build and manage AI solutions; third parties' use and adoption of our products and technologies, and the benefits and impacts thereof; researchers being able to generate molecules that are optimized according to scientists' specific needs by leveraging BioNeMo microservices; generative AI allowing third parties to build sophisticated models and seamlessly integrate AI into the antibody design process and to create the next generation of medicines that will bring the most value to patients; generative AI changing the future of patient care; and third parties' ability to mitigate widespread staffing shortages and increase access to high-quality care while improving outcomes for patients with generative AI are forward-looking statements that are subject to risks and uncertainties that could cause results to be materially different than expectations. Important factors that could cause actual results to differ materially include: global economic conditions; our reliance on third parties to manufacture, assemble, package and test our products; the impact of technological development and competition; development of new products and technologies or enhancements to our existing product and technologies; market acceptance of our products or our partners' products; design, manufacturing or software defects; changes in consumer preferences or demands; changes in industry standards and interfaces; unexpected loss of

performance of our products or technologies when integrated into systems; as well as other factors detailed from time to time in the most recent reports NVIDIA files with the Securities and Exchange Commission, or SEC, including, but not limited to, its annual report on Form 10-K and quarterly reports on Form 10-Q. Copies of reports filed with the SEC are posted on the company's website and are available from NVIDIA without charge. These forward-looking statements are not guarantees of future performance and speak only as of the date hereof, and, except as required by law, NVIDIA disclaims any obligation to update these forward-looking statements to reflect future events or circumstances.

Janette Ciborowski
+1-734-330-8817
jciborowski@nvidia.com