

NVIDIA Unveils Large Language Models and Generative AI Service to Advance Life Sciences R&D

Part of NVIDIA AI Foundations, New BioNeMo Cloud Service Accelerates Life Sciences Research, Drug Discovery and Protein Engineering; Amgen and a Dozen Startups Among Early Access Customers

GTC—NVIDIA today announced an expanded set of generative AI cloud services for customizing AI foundation models to accelerate the creation of new proteins and therapeutics, as well as research in the fields of genomics, chemistry, biology and molecular dynamics.

Part of [NVIDIA AI Foundations](#), the new [BioNeMo™ Cloud service](#) offering — for both AI model training and inference — accelerates the most time-consuming and costly stages of drug discovery. It enables researchers to fine-tune generative AI applications on their own proprietary data, and to run AI model inference directly in a web browser or through new cloud application programming interfaces (APIs) that easily integrate into existing applications.

“The transformative power of generative AI holds enormous promise for the life science and pharmaceutical industries,” said Kimberly Powell, vice president of healthcare at NVIDIA. “NVIDIA’s long collaboration with pioneers in the field has led to the development of BioNeMo Cloud Service, which is already serving as an AI drug discovery laboratory. It provides pretrained models and allows customization of models with proprietary data that serve every stage of the drug-discovery pipeline, helping researchers identify the right target, design molecules and proteins, and predict their interactions in the body to develop the best drug candidate.”

Amgen Among Early Users

Amgen, one of the world’s leading biotechnology companies, is already using the service to advance its research and development efforts.

“BioNeMo is dramatically accelerating our approach to biologics discovery,” said Peter Grandsard, executive director of Biologics Therapeutic Discovery, Center for Research Acceleration by Digital Innovation at Amgen. “With it, we can pretrain [large language models](#) for molecular biology on Amgen’s proprietary data, enabling us to explore and develop therapeutic proteins for the next generation of medicine that will help patients.”

Generative AI Supercharges Drug Discovery Pipeline

BioNeMo Cloud service includes [pretrained AI models](#) to help researchers build AI pipelines for drug development. It has been adopted by drug-discovery companies including Evozyne and Insilico Medicine to support data-driven drug design for new therapeutic candidates.

Generative AI models can rapidly identify potential drug molecules — in some cases designing compounds or protein-based therapeutics from scratch. Trained on large-scale datasets of small molecules, proteins, DNA and RNA sequences, these models can predict the 3D structure of a protein and how well a molecule will dock with a target protein.

New Generative AI Models Available With BioNeMo Service Early Access

BioNeMo now has [six new optimized, open-source models](#), in addition to its previously announced MegaMolBART generative chemistry model, ESM1nv protein language model and OpenFold protein structure prediction model. They include:

- **AlphaFold2:** A deep learning model that reduces the time it takes to determine a protein’s structure from years to minutes or even seconds, just by using its amino acid sequence, developed by [DeepMind](#) and already used by over a million researchers.
- **DiffDock:** To help researchers understand how a drug molecule will bind with a target protein, this model predicts the 3D orientation and docking interaction of small molecules with high accuracy and computational efficiency.
- **ESMFold:** This protein structure prediction model, using [Meta AI](#)’s **ESM2** protein language model, can estimate the 3D structure of a protein based on a single amino acid sequence, without requiring examples of several similar sequences.
- **ESM2:** This protein language model is used for inferring machine representations of proteins which are useful for downstream tasks such as protein structure prediction, property prediction and molecular docking.
- **MoFlow:** Used for molecular optimization and small molecule generation, this generative chemistry model creates molecules from scratch, coming up with diverse chemical structures for potential therapeutics.
- **ProtGPT-2:** This language model generates novel protein sequences to help researchers design proteins with unique structures, properties and functions.

The BioNeMo Service makes these generative AI models easily accessible through a browser-based interface for interactive inference and protein structure visualization. And by pairing BioNeMo with the supercomputing resources in [NVIDIA DGX™](#)

[Cloud](#), researchers can customize their models on a fully managed software service using [NVIDIA Base Command™ Platform](#) and the [NVIDIA AI Enterprise](#) software suite.

Pharma Companies, Startups Tap BioNeMo to Optimize AI Workflows

Pharmaceutical companies and drug discovery startups are using BioNeMo today and, in many cases, seeing significant results.

Amgen pretrained and fine-tuned BioNeMo's ESM model architecture using its own proprietary data on antibodies. It was able to slash the time it takes to train five custom models for molecule screening and optimization from three months to a few weeks on DGX Cloud.

Researchers at Evozyne, a Chicago-based biotechnology company and member of the [NVIDIA Inception](#) program for cutting-edge startups, have collaborated with NVIDIA to develop a [BioNeMo-based deep learning model](#) called the Protein Transformer Variational AutoEncoder. The generative AI model, fine-tuned on Evozyne's proprietary protein data, enables the design of synthetic variants with significantly improved performance compared to enzymes found in nature.

Insilico Medicine, a premier member of NVIDIA Inception, is using BioNeMo to accelerate the early drug discovery process, which traditionally takes more than four years and costs around \$500 million. Using generative AI from end to end, Insilico was able to [identify a preclinical candidate](#) drug in one-third of the time and for one-tenth of the cost. The drug is expected to soon enter phase 2 clinical trials with patients.

Sign up for [early access to BioNeMo](#).

Discover the latest in [AI and healthcare](#) — including three sessions with speakers from Amgen, a session from Evozyne and another from DeepMind on AlphaFold — at GTC, running online through Thursday, March 23. [Registration](#) is free.

Watch NVIDIA founder and CEO Jensen Huang discuss the [BioNeMo Cloud service in his GTC keynote](#) on demand.

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 the creation of the metaverse. NVIDIA is now a full-stack computing 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 our collaborations with Amgen, Evozyne and Insilico Medicine; the benefits, impact, performance, features and availability of our products and technologies, including NVIDIA AI Foundations such as the new BioNeMo Cloud service offering, BioNeMo models including the MegaMolBART generative chemistry model, ESM1nv protein language model, OpenFold protein structure prediction model, AlphaFold2, DiffDock, ESMFold, ESM2, MoFlow and ProtGPT-2, NVIDIA DGX Cloud, NVIDIA Base Command Platform and the NVIDIA AI Enterprise software suite; the transformative power of generative AI holding enormous promise for the life science and pharmaceutical industries; pharmaceutical companies and drug discovery startups using BioNeMo today and, in many cases, seeing significant results; and the preclinical candidate drug soon entering phase 2 clinical trials with patients 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.

© 2023 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, BioNeMo, DGX Cloud and NVIDIA Base Command are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated. Features, pricing, availability, and specifications are subject to change without notice.

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