

NVIDIA and Global Partners Launch NIM Agent Blueprints for Enterprises to Make Their Own AI

- *Catalog of Customizable Workflows Speeds Deployments of Core Generative AI Use Cases, Starting With Customer Service, Drug Discovery and Data Extraction for PDFs, With More to Come*
- *Companies Can Build and Operationalize Their AI Applications — Creating Data-Driven AI Flywheels — Using NIM Agent Blueprints Along With NIM Microservices and NeMo Framework, All Part of the NVIDIA AI Enterprise Platform*
- *Accenture, Cisco, Dell Technologies, Deloitte, Hewlett Packard Enterprise, Lenovo, SoftServe, World Wide Technology Among First Partners Delivering NIM Agent Blueprints to World's Enterprises*

NVIDIA today announced [NVIDIA NIM™ Agent Blueprints](#), a catalog of pretrained, customizable AI workflows that equip millions of enterprise developers with a full suite of software for building and deploying generative AI applications for canonical use cases, such as customer service avatars, retrieval-augmented generation and drug discovery virtual screening.

NIM Agent Blueprints provide a jump-start for developers creating AI applications that use one or more AI agents. They include sample applications built with [NVIDIA NeMo™](#), [NVIDIA NIM](#) and partner microservices, reference code, customization documentation and a Helm chart for deployment.

Enterprises can modify NIM Agent Blueprints using their business data and run their generative AI applications across accelerated data centers and clouds. With NIM Agent Blueprints, enterprises can continually refine their AI applications based on user feedback, creating a data-driven AI flywheel.

The first NIM Agent Blueprints now available include a digital human workflow for customer service, a generative virtual screening workflow for computer-aided drug discovery and a multimodal PDF data extraction workflow for enterprise retrieval-augmented generation (RAG) that can use vast quantities of business data for more accurate responses. NIM Agent Blueprints are free for developers to experience and download and can be deployed in production with the [NVIDIA AI Enterprise](#) software platform.

Global system integrators and technology solutions providers [Accenture](#), [Deloitte](#), [SoftServe](#) and [World Wide Technology](#) (WWT) are bringing NVIDIA NIM Agent Blueprints to enterprises worldwide. [Cisco](#), [Dell Technologies](#), [Hewlett Packard Enterprise](#) and [Lenovo](#) are offering full-stack NVIDIA-accelerated infrastructure and solutions to speed NIM Agent Blueprints deployments.

“Generative AI is advancing at lightspeed. Frontier model capabilities are growing exponentially with a continuous stream of new applications,” said Jensen Huang, founder and CEO of NVIDIA. “The enterprise AI wave is here. With the NVIDIA AI Enterprise toolkit — including NeMo, NIM microservices and the latest NIM Agent Blueprints — our expansive partner ecosystem is poised to help enterprises customize open-source models, build bespoke AI applications and deploy them seamlessly across any cloud, on premises or at the edge.”

Digital Human NIM Agent Blueprint Advances Customer Service

[Gartner®](#) reports that 80% of conversational offerings will embed generative AI by 2025, up from 20% in 2024⁽¹⁾. The digital human NIM Agent Blueprint for customer service helps enterprises rapidly prepare for this coming shift, bringing enterprise applications to life with a 3D animated avatar interface.

With approachable, humanlike interactions, customer service applications can provide more engaging user experiences compared to traditional customer service options. Powered by [NVIDIA Tokkio](#) technologies, the digital human workflow features NVIDIA software including [NVIDIA ACE](#), [NVIDIA Omniverse RTX™](#), [NVIDIA Audio2Face™](#) and [Llama 3.1](#) NIM microservices, and is designed to integrate with existing enterprise generative AI applications built using RAG.

Multimodal PDF Data Extraction NIM Agent Blueprint Taps Business Data

The multimodal PDF data extraction workflow for enterprise RAG uses [NVIDIA NeMo Retriever](#) NIM microservices to unlock insights from massive volumes of enterprise PDF data. With this workflow, developers can create digital humans, AI agents or customer service chatbots that can quickly become experts on any topic captured within their corpus of PDF data.

Using the workflow, enterprises can combine NeMo Retriever NIM microservices with community or custom models to build high-accuracy, multimodal retrieval pipelines that can be deployed wherever enterprise data resides.

Generative Virtual Screening NIM Agent Blueprint Accelerates Drug Discovery

The generative virtual screening NVIDIA NIM Agent Blueprint for drug discovery accelerates the identification and optimization of promising drug-like molecules, significantly reducing time and cost by generating molecules with favorable properties and higher probabilities of success.

Researchers and application developers can quickly customize and deploy AI models for 3D protein structure prediction, small molecule generation and molecular docking. This blueprint incorporates [NVIDIA NIM microservices](#) — including AlphaFold2, [MolMIM](#) and [DiffDock](#) — to accelerate the virtual screening of small molecules using generative models.

In combination with other tools available in [NVIDIA BioNeMo™](#), enterprises can easily connect multiple NIM Agent Blueprints to build increasingly sophisticated AI applications and accelerate their drug discovery work.

Additional blueprints will be released monthly for workflows to build generative AI applications for customer experience, content generation, software engineering, and product research and development.

NVIDIA Partner Ecosystem Amplifies Enterprise Generative AI Success

NVIDIA partners are readying to help the world's enterprises rapidly build and deploy their own generative AI applications using NIM Agent Blueprints.

Global professional services company Accenture will add NVIDIA NIM Agent Blueprints to its [Accenture AI Refinery™](#), unveiled last month.

"Across industries, generative AI is acting as a catalyst for companies looking to reinvent with tech, data and AI," said Julie Sweet, chair and CEO of Accenture. "By integrating NVIDIA's catalog of workflows into Accenture's AI Refinery, we can help our clients develop custom AI systems at speed and reimagine how they do business and serve their customers to drive stronger business outcomes and create new value."

Global consulting firm Deloitte will integrate NVIDIA NIM Agent Blueprints into its deep portfolio of NVIDIA-powered solutions.

"While many organizations are still working to fully harness the potential of GenAI, its implementation is steadily enhancing efficiencies and productivity," said Jason Girzadas, CEO of Deloitte US. "By embedding NVIDIA's NIM Agent Blueprints into enterprise solutions that are built on NVIDIA NIM microservices, Deloitte is engaging with our clients to innovate faster, unlock new growth opportunities and define AI-competitive advantage."

IT consulting and digital services provider SoftServe is integrating NIM Agent Blueprints into its generative AI portfolio to speed enterprise adoption.

"Every enterprise knows generative AI is central to modernizing their operations, but not every enterprise knows where to begin their generative AI journey," said Harry Propper, CEO of SoftServe. "Adding NVIDIA NIM Agent Blueprints into the SoftServe Gen AI Solutions portfolio gives clients proven frameworks for developing AI applications that put their own data to work."

A solution provider for the majority of Fortune 100 companies, WWT will assist enterprises in building NIM Agent Blueprints that leverage their business data.

"WWT is committed to helping enterprises harness the power of AI as a catalyst for business transformation," said Jim Kavanaugh, cofounder and CEO of World Wide Technology. "WWT's AI Proving Ground, equipped with NVIDIA NIM Agent Blueprints and coupled with our data scientists, consultants and high-performance architecture engineers, offers a comprehensive resource for our clients to experiment with, validate and scale AI solutions."

Enterprises can develop and deploy NIM Agent Blueprints on NVIDIA AI platforms with compute, networking and software provided by NVIDIA's global server manufacturing partners.

These include Cisco Nexus HyperFabric AI clusters with NVIDIA, the Dell AI Factory with NVIDIA, NVIDIA AI Computing by HPE and HPE Private Cloud AI, as well as Lenovo Hybrid AI solutions powered by NVIDIA.

"Cisco, together with NVIDIA, created a revolutionary, flexible and simple-to-deploy AI infrastructure with Nexus HyperFabric," said Chuck Robbins, chair and CEO of Cisco. "Combining Cisco innovation with NVIDIA NIM Agent Blueprints offers customers a simple and secure way to deploy generative AI fast and efficiently, with the adaptability they need to build and customize new applications at scale."

"Dell Technologies and NVIDIA are making it easy for enterprises to unlock the power of AI-enabled applications and agents," said Michael Dell, founder and CEO of Dell Technologies. "Incorporating NVIDIA NIM Agent Blueprints into the Dell AI Factory with NVIDIA provides an express lane for the transformative value of AI."

"HPE and NVIDIA are expanding on our recent blockbuster collaboration to deliver NVIDIA AI Computing by HPE," said Antonio Neri, president and CEO of Hewlett Packard Enterprise. "By integrating NVIDIA NIM Agent Blueprints into our co-developed turnkey HPE Private Cloud AI solution, we will enable enterprises to focus resources on developing new AI use cases that boost productivity and unlock new revenue streams."

"Generative AI is a full-stack challenge that requires accelerated infrastructure, specialized software and services, and powerful AI-ready devices that can maximize the capabilities of Hybrid AI," said Yuanqing Yang, chairman and CEO of

Lenovo. “NVIDIA NIM Agent Blueprints, combined with Lenovo’s comprehensive, end-to-end portfolio, give enterprises a head start for building generative AI applications that they can run everywhere on Lenovo Hybrid AI.”

Enterprises can experience [NVIDIA NIM Agent Blueprints](#) today.

About NVIDIA

[NVIDIA](#) (NASDAQ: NVDA) is the world leader in accelerated computing.

(1) Gartner, “Emerging Tech: Navigating the Hurdles of Digital Humans” by Evan Brown, 6 May 2024

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 NIM Agent Blueprints, NVIDIA NeMo, NVIDIA NIM, NVIDIA AI Enterprise software platform, NVIDIA Tokkio, NVIDIA ACE, NVIDIA Omniverse RTX, NVIDIA Audio2Face, NVIDIA NeMo Retriever, and NVIDIA BioNeMo NIM microservices — including AlphaFold2, MolMIM and DiffDock; third parties using or adopting NVIDIA products, technologies and platforms, and the benefits and impacts thereof; our collaboration with third parties and the benefits and impacts thereof; generative AI advancing at lightspeed; frontier model capabilities growing exponentially with a continuous stream of new applications; the enterprise AI wave being here with the NVIDIA AI Enterprise toolkit – including NeMo, NIM microservices and the latest NIM Agent Blueprints – NVIDIA’s expansive partner ecosystem being poised to help enterprises customize open-source models, build bespoke AI applications, and deploy them seamlessly across any cloud, on premises or at the edge; with approachable, humanlike interactions, customer service applications being able to provide more engaging user experiences compared to traditional customer service options; NVIDIA partners readying to help the world’s enterprises rapidly build and deploy their own generative AI applications using NIM Agent Blueprints; generative AI acting as a catalyst for companies looking to reinvent with tech, data and AI; GenAI’s implementation steadily enhancing efficiencies and productivity; and generative AI being a full-stack challenge that requires accelerated infrastructure, specialized software and services, and powerful AI-ready devices that can maximize the capabilities of Hybrid 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.

Many of the products and features described herein remain in various stages and will be offered on a when-and-if-available basis. The statements above are not intended to be, and should not be interpreted as a commitment, promise, or legal obligation, and the development, release, and timing of any features or functionalities described for our products is subject to change and remains at the sole discretion of NVIDIA. NVIDIA will have no liability for failure to deliver or delay in the delivery of any of the products, features or functions set forth herein.

© 2024 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, Audio2Face, BioNeMo, NVIDIA NeMo, NVIDIA NIM and NVIDIA Omniverse RTX are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. 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.

Shannon McPhee
NVIDIA Corporation
+1-310-920-9642
smcphee@nvidia.com