

---

# NVIDIA NIM Blueprints

---

STEVE MCDOWELL, CHIEF ANALYST  
9/4/24

---

## CONTEXT

---

NVIDIA launched its new [NIM Agent Blueprints](#), a catalog of pre-trained, customizable AI workflows to help enterprise developers quickly build and deploy generative AI applications for critical use cases, such as customer service, drug discovery, and data extraction from PDFs.

The blueprints provide sample applications, reference code, and deployment tools that can be customized with business data and run across accelerated data centers or clouds.

The first set of NIM Agent Blueprints includes:

- **Digital Human Blueprint for customer service**, creating 3D avatars for enhanced user interaction.
- **Multimodal PDF Data Extraction Blueprint**, enabling AI agents to extract insights from large volumes of business data.
- **Generative Virtual Screening Blueprint for drug discovery**, accelerating the identification of promising drug candidates.

NVIDIA's partner ecosystem, including Accenture, Cisco, Dell Technologies, and others, will deliver these blueprints to enterprises globally, speeding up AI adoption.

The new solutions, built on NVIDIA's AI Enterprise Platform using tools like NeMo and NIM microservices, allow enterprises to continually refine their AI applications, driving innovation and business transformation.

---

## NEW: NIM BLUEPRINTS

---

NVIDIA NIM Blueprints help enterprises quickly build and operationalize AI-driven applications using a set of prebuilt workflows. These blueprints provide developers with tools to create customized AI systems to fit specific business needs.

- **Pretrained AI Workflows:** Each NIM Blueprint includes ready-made, customizable AI workflows tailored to specific use cases, drastically reducing the time and complexity of building AI applications from scratch.
- **Integration with Business Data:** Enterprises can modify NIM Agent Blueprints using their business data, allowing for a more personalized and relevant AI application deployment.

---

## KEY USE CASES

---

NIM Agent Blueprints offer targeted solutions for some of the most critical generative AI use cases in today's business environment:

- **Customer Service Automation:**
  - The **Digital Human Blueprint** provides tools for creating 3D animated avatars to serve as AI-driven customer service agents.
  - This blueprint enhances user interaction with lifelike, conversational AI agents, improving customer engagement.
- **Drug Discovery:**
  - The **Generative Virtual Screening Blueprint** accelerates identifying and optimizing drug molecules through AI-driven virtual screening.
  - It integrates NVIDIA NIM microservices like **AlphaFold2** and **MolMIM** to predict 3D protein structures and perform molecular docking.
- **PDF Data Extraction:**
  - The **Multimodal PDF Data Extraction Blueprint** enables enterprises to unlock valuable insights from vast amounts of PDF data.
  - AI agents using this workflow can extract, analyze, and retrieve information from business documents to improve decision-making processes.

---

## BUILT ON THE NVIDIA AI ENTERPRISE PLATFORM

---

NIM Blueprints leverage NVIDIA's AI infrastructure, integrating seamlessly with the NVIDIA AI Enterprise Platform:

- **NeMo and NIM Microservices:** NIM Blueprints rely on **NeMo** for LLMs and **NIM microservices** for infrastructure and AI service orchestration. These tools allow for advanced customization and optimization.
- **Ease of Deployment:** NIM Blueprints include sample applications, reference code, customization instructions, and Helm charts, allowing for rapid deployment across cloud or on-premise environments.

---

## CONTINUOUS AI FLYWHEEL

---

One of the key advantages of NIM Blueprints is their ability to drive continuous AI improvement:

- **Data-Driven Refinement:** Enterprises can feed user feedback and real-time business data into their AI models, creating a feedback loop (or "AI flywheel") that continuously enhances the accuracy, efficiency, and relevance of the AI application.
- **Adaptability:** As the AI flywheel turns, these models improve over time, making the AI applications more responsive to changing business needs and external factors.

---

## PARTNER ECOSYSTEM

---

NVIDIA has partnered with major global technology leaders to support the deployment and scalability of NIM Blueprints, ensuring that enterprises have access to world-class infrastructure and services:

- **Accenture, Deloitte, and SoftServe:** These global consulting firms help enterprises integrate NIM Blueprints into their operations, accelerating AI adoption across industries.
- **Technology Infrastructure Partners:** Companies like **Cisco, Dell Technologies, Hewlett Packard Enterprise (HPE), and Lenovo** provide the hardware and cloud infrastructure to power and deploy these AI solutions.

---

## FEATURES AND BENEFITS

---

NVIDIA NIM Blueprints come with several features that make them attractive for enterprises looking to quickly integrate AI into their business processes:

- **Prebuilt Solutions:** Each blueprint includes sample applications, reference code, and detailed documentation to simplify development and customization.
- **Full Flexibility:** The blueprints are scalable and deployable across any infrastructure, whether in the cloud, on-premises or at the edge.
- **Customizability:** Developers can modify the blueprints using specific business data and tailor them to fit particular organizational workflows or processes.
- **Free to Download:** NIM Blueprints are free for developers to download and experiment with, reducing the cost barrier for AI exploration and development.

---

## FUTURE EXPANSION

---

NVIDIA plans to expand the NIM Blueprints catalog with additional workflows designed to address more generative AI applications, such as:

- **Content Generation**
- **Software Engineering**
- **Product Research and Development**
- **Customer Experience Optimization**

---

## BACKGROUND: WHAT IS NVIDIA NIM?

---

NVIDIA NIM Blueprints are built using NVIDIA NeMo and NIM platforms. These are key components of its broader AI enterprise strategy to help businesses efficiently create, customize, and deploy large-scale generative AI models.

The platforms, integral to NVIDIA's AI Enterprise ecosystem, provide the tools and infrastructure necessary to develop advanced AI-driven applications across industries.

---

## NVIDIA NEMO

---

NVIDIA NeMo is an open-source framework designed to help developers build and fine-tune LLMs and other generative AI models. It offers several features that make it essential for enterprises looking to integrate LLMs into their workflows:

- **Model Library:** NeMo comes with a rich collection of pre-trained models for various tasks, including automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and question-answering (QA). Developers can use these models as a starting point for building specialized AI systems.
- **Fine-Tuning and Customization:** NeMo allows for fine-tuning models using organization-specific data, enhancing the accuracy and relevance of AI applications.
- **Scalability:** With NeMo, enterprises can scale their AI models across various infrastructure environments, from cloud to on-premises, leveraging NVIDIA's GPU-accelerated infrastructure for high performance.
- **NeMo Guardrails:** An important aspect of NeMo is Guardrails, a set of tools designed to ensure the safe, secure, and responsible use of AI models. This includes setting boundaries on model outputs to maintain alignment with business objectives and regulatory requirements.

---

## NVIDIA NIM

---

NVIDIA NIM is NVIDIA's comprehensive platform that provides microservices and workflows for developing AI applications involving AI agents. It focuses on enabling enterprises to operationalize AI solutions across multiple use cases, from customer service automation to drug discovery.

- **Pretrained Workflows (NIM Blueprints):** NIM comes with pre-trained AI workflows, known as NIM Blueprints, which offer customizable templates for various applications. These blueprints simplify the development of AI solutions by providing sample applications, reference code, and documentation.
- **Modularity and Customization:** Enterprises can modify NIM Blueprints with their business data, creating bespoke AI solutions tailored to specific organizational needs. This flexibility allows for rapidly deploying

generative AI models across different environments, including cloud, on-premises, and edge deployments.

- **Microservices Integration:** NIM includes a set of microservices that handle core AI functionalities, such as AI model management, data processing, and integration with other enterprise systems. This modular approach makes scaling and adapting AI solutions easier as business needs evolve.
- **AI Flywheel:** One of NIM's most powerful aspects is its ability to create a data-driven AI flywheel. This flywheel allows organizations to continuously improve AI models by incorporating real-time data and user feedback. This feedback loop ensures that AI applications become more accurate and responsive.

## ANALYSIS

---

NVIDIA's differentiation in AI and the secret behind much of its success is the company's software-first approach. There's no dearth of silicon-based AI accelerators, but there is a paucity of enabling software solutions. NVIDIA's Enterprise AI offerings keep the company at the forefront of mainstreaming AI. Its new NIM Blueprints are part of this effort.

NVIDIA's NIM Blueprints are a major step forward in the enterprise AI space. By offering customizable, pre-trained workflows that can be easily tailored and deployed, NVIDIA helps businesses overcome many traditional challenges associated with AI development and adoption. Its strong ecosystem of partners, coupled with a focus on critical use cases and continuous innovation, positions NVIDIA as a leading force in the generative AI revolution.

The current wave of generative AI innovation is transforming industries at an unprecedented pace, and companies that don't capitalize on this momentum risk falling behind. By providing enterprises with ready-made tools for generative AI, NVIDIA enables businesses to adopt these technologies at scale and speed.

This becomes critical as more businesses realize AI's transformative potential but often struggle with where and how to start. With NIM Blueprints, NVIDIA has essentially created a blueprint for AI success, offering businesses a clear and actionable path to follow.

With this new offering, NVIDIA continues to cement its role as a dominant player in the enterprise AI market, and it is clear that the company's AI strategy is both visionary and deeply aligned with the needs of modern businesses.



© Copyright NAND Research.

NAND Research is a registered trademark of NAND Research LLC, All Rights Reserved.

This document may not be reproduced, distributed, or modified, in physical or electronic form, without the express written consent of NAND Research. Questions about licensing or use of this document should be directed to [info@nand-research.com](mailto:info@nand-research.com).

The information contained within this document was believed by NAND Research to be reliable and is provided for informational purposes only. The content may contain technical inaccuracies, omissions, or typographical errors. This document reflects the opinions of NAND Research, which is subject to change. NAND Research does not warranty or otherwise guarantee the accuracy of the information contained within.

NAND Research is a technology-focused industry analyst firm providing research, customer content, market and competitive intelligence, and custom deliverables to technology vendors, investors, and end-customer IT organizations.

Contact NAND Research via email at [info@nand-research.com](mailto:info@nand-research.com) or visit our website at [nand-research.com](http://nand-research.com).