# Nvidia launches Nemotron models to advance agentic artificial intelligence



Nvidia has made significant strides in the development of agentic artificial intelligence (AI) with the launch of its Nemotron model family, unveiled during a recent announcement. The move showcases Nvidia's commitment to expanding its role beyond hardware into the burgeoning field of agentic AI, which emphasises autonomous, interactive digital agents that can work alongside humans.

The Nemotron models, trained using techniques from Meta's Llama datasets, consist of three variants: Nano, Super, and Ultra. Each model is optimised for different applications, catering to various computing needs. The Nano variant is designed for efficient low-latency applications on edge devices, employing 4 billion parameters; the Super version, with 49 billion parameters, focuses on high accuracy and throughput; while the Ultra configuration, featuring 253 billion parameters, offers the utmost precision tailored for data centre environments.

Rev Lebaredian, Nvidia's Vice President of Omniverse and simulation technology, explained in a briefing that these models are specifically tailored for critical agentic tasks such as instruction following, chats, function calling, coding, and mathematical operations. He noted, “AI agents are the digital workforce that will work for us and work with us, and so the Nemotron model family is for agentic AI.”

Additionally, Nvidia has introduced AI orchestration blueprints, further positioning itself as a leader in the growing agentic AI ecosystem. These blueprints serve as frameworks for managing multiple AI agents across various systems, enabling them to work harmoniously within enterprises. To this end, Nvidia has partnered with several orchestration companies, including LangChain, LlamaIndex, CrewAI, Daily, and Weights & Biases, to develop specific blueprints tailored to various tasks, such as navigating code repositories or generating structured reports from internet searches.

In a related development, Nvidia showcased its enhanced AI Enterprise software platform at CES 2025, indicating its increasing relevance to independent software vendors (ISVs) like Microsoft, SAP, Salesforce, and ServiceNow. The platform will facilitate the deployment of agentic AI solutions, as multiple major ISVs have already begun leveraging Nvidia’s capabilities. In particular, Zoho has strengthened its ties with Nvidia, underscoring that the tech giant is at the forefront of the autonomous AI agent movement.

Highlighting the importance of orchestration, Lebaredian stated, “Making multiple agents work together smoothly is key to deploying agentic AI.” Nvidia’s orchestration layer will enable enterprises to coordinate numerous AI agents effectively, ensuring they perform as intended. The orchestration layer aims to provide a cohesive framework, allowing agents from different partners to work together seamlessly, thereby enhancing the flexibility of AI implementations across various sectors.

As part of its forward-looking strategy, Nvidia is also partnering with consulting giant Accenture to establish an “AI Refinery for Industry,” built on the Nvidia AI Enterprise platform. This initiative aims to streamline the deployment of AI agents across diverse industries such as automotive, consumer goods manufacturing, public sector, and technology. The project intends to produce over 100 industry-specific AI solutions by 2026, facilitating rapid development and implementation.

Nvidia’s advancements also encompass the introduction of a family of open large language models (LLMs), called Llama Nemotron, which are planned to serve as a fundamental framework for agentic AI applications. These models are designed to optimise various workloads, including customer support, fraud detection, and supply chain management. Additionally, the Cosmos Nemotron vision language models provide visual understanding capabilities, pivotal for robotics applications and other physics-aware agentic AI innovations.

As Nvidia continues to lead in agentic AI development, it reflects a broader industry trend towards enhancing business operations through advanced artificial intelligence solutions. In this dynamic landscape, enterprises are increasingly exploring innovative applications and infrastructure necessities to fully harness the potential of autonomous agents in their workflows.

Source: [Noah Wire Services](https://www.noahwire.com)

## References

* <https://www.constellationr.com/blog-news/insights/nvidia-moves-advance-agentic-ai-use-cases-ces-2025> - Corroborates Nvidia's launch of the Nemotron model family, the different variants (Nano, Super, Ultra), and their applications. Also mentions the introduction of AI orchestration blueprints and partnerships with various companies.
* <https://www.constellationr.com/blog-news/insights/nvidia-moves-advance-agentic-ai-use-cases-ces-2025> - Supports the information about Nvidia's AI Enterprise software platform, its relevance to ISVs, and the partnership with Accenture for the 'AI Refinery for Industry'.
* <https://www.constellationr.com/blog-news/insights/nvidia-moves-advance-agentic-ai-use-cases-ces-2025> - Details the use of Nemotron models for various agentic tasks such as instruction following, chats, function calling, coding, and mathematical operations.
* <https://the-decoder.com/nvidia-improves-metas-llama-model-with-new-training-approach/> - Explains the training techniques used for the Llama Nemotron models, including the use of Meta's Llama datasets and the creation of new datasets like HelpSteer2 and HelpSteer2-Preference.
* <https://www.constellationr.com/blog-news/insights/nvidia-moves-advance-agentic-ai-use-cases-ces-2025> - Mentions the introduction of AI orchestration blueprints and partnerships with companies like LangChain, LlamaIndex, CrewAI, Daily, and Weights & Biases.
* <https://www.constellationr.com/blog-news/insights/nvidia-moves-advance-agentic-ai-use-cases-ces-2025> - Highlights the importance of orchestration in deploying agentic AI and Nvidia's efforts to enable multiple AI agents to work together smoothly.
* <https://www.constellationr.com/blog-news/insights/nvidia-moves-advance-agentic-ai-use-cases-ces-2025> - Details the partnership with Accenture to establish an 'AI Refinery for Industry' and the goal to produce over 100 industry-specific AI solutions by 2026.
* <https://www.constellationr.com/blog-news/insights/nvidia-moves-advance-agentic-ai-use-cases-ces-2025> - Corroborates the introduction of the Llama Nemotron family for agentic AI applications, including customer support, fraud detection, and supply chain management.
* <https://www.constellationr.com/blog-news/insights/nvidia-moves-advance-agentic-ai-use-cases-ces-2025> - Mentions the Cosmos Nemotron vision language models and their visual understanding capabilities for robotics and other physics-aware agentic AI innovations.
* <https://www.constellationr.com/blog-news/insights/nvidia-moves-advance-agentic-ai-use-cases-ces-2025> - Reflects the broader industry trend towards enhancing business operations through advanced artificial intelligence solutions and the increasing exploration of autonomous agents in workflows.