



Japan's Enterprises and Startups Build Industry-Specialized AI With NVIDIA Nemotron Open Models

News Summary:

- Institution of Science Tokyo, SoftBank Corp., SB Intuitions Corp. and Stockmark are adopting NVIDIA Nemotron to build locally developed AI models designed to serve Japanese users, businesses and institutions amid the country's demographic and workforce transition.
- Japanese enterprises avatarin, ENEOS Holdings, Hitachi and NTT DATA are building Japanese-language AI applications with NVIDIA Nemotron, from remote-presence robotics to enterprise agents and specialized medical and contact centers.
- Sakana AI is integrating NVIDIA Nemotron into its Fugu model-routing platform, expanding the set of AI models Fugu can intelligently orchestrate to dynamically select the best model for each task.

NVIDIA today announced that leading Japanese enterprises, startups and research institutions are building industry-specialized AI models and applications with [NVIDIA Nemotron™](#) open models, data and libraries, accelerating the development of AI tailored to Japan's language, industries and workforce.

[Open models](#) are the foundation of national AI ecosystems, giving organizations the ability to customize, deploy and govern AI they control.

In Japan, these capabilities are increasingly important as the country addresses an aging population and workforce transition, driving demand for AI tailored to local industries that helps strengthen the workforce, sustain productivity and accelerate innovation.

"Every nation and every company should own and control its intelligence infrastructure. Open models make that possible," said Jensen Huang, founder and CEO of NVIDIA. "They give countries, enterprises and researchers the freedom to inspect, improve, adapt, secure and deploy AI for their own needs. Together with Japan's AI leaders, we are advancing an open AI ecosystem that accelerates discovery, strengthens national capability and ensures every society can participate in -- and benefit from -- the AI revolution."

Building Specialized AI for Japan With NVIDIA Nemotron

Across Japan, developers are building specialized AI with NVIDIA Nemotron open models and datasets, tailoring them to the country's industries and public-sector needs.

Institute of Science Tokyo developed its Swallow family of open foundation models using NVIDIA Nemotron datasets and the NVIDIA NeMo™ software stack for continual pretraining and post-training. Swallow models enhance Japanese language and reasoning performance while preserving the underlying models' core English, math and coding capabilities. Enterprises are customizing and deploying Swallow for specialized use cases, including financial-document translation and asset-management report generation.

SB Intuitions Corp., SoftBank Corp.'s generative AI research subsidiary, trained its Sarashina series of homegrown generative AI models using NVIDIA Nemotron, including the NVIDIA NeMo RL and Megatron-LM libraries. Sarashina3 mini has been selected by Japan's Digital Agency for use in specialized AI use cases. SoftBank Corp. has also developed and deployed a Large Telecom Model (LTM), using both Sarashina and NVIDIA Nemotron, to enable autonomous telecom network operations.

[Stockmark](#)'s specialized Japanese-language document-understanding model, released today, is based on the NVIDIA Nemotron 3 Nano Omni model. The company is also developing enterprise knowledge applications using NVIDIA NeMo Retriever™ and the [Nemotron-Personas-Japan](#) dataset, serving customers across Japan's manufacturing, energy and chemical industries through Japan's Generative AI Accelerator Challenge national project.

Transforming Japan's Industries With NVIDIA Nemotron

Japanese enterprises are using NVIDIA Nemotron to modernize essential services, improve productivity and support the country's workforce.

AI and robotics startup avatarin is using NVIDIA Nemotron open models and NVIDIA NeMo to develop Japanese-language speech and reasoning capabilities for enterprise AI agents. [NVIDIA HGX™](#) B300 systems provide the private AI infrastructure that enables those agents to securely analyze customer conversations and access enterprise knowledge for more accurate responses, while [NVIDIA Jetson™](#) powers edge AI capabilities, including digital avatar systems being deployed at airports and other locations across Japan.

ENEOS Holdings is using NVIDIA Nemotron open models with the [NVIDIA AI-Q Blueprint](#) and [NVIDIA ALCHEMI NIM](#) microservices to advance agentic AI workflows for energy and materials R&D. Researchers are using these technologies to integrate technical document search, vision and language understanding, and simulation-backed molecular screening, helping accelerate materials exploration for applications such as immersion-cooling fluids and advanced catalysts.

NTT DATA, an operating subsidiary of NTT, used NVIDIA Nemotron-Personas-Japan to augment training data for its proprietary tsuzumi 2 model, improving question-answering accuracy and enhancing responses to questions requiring additional knowledge. The company is also looking to deploy a scalable multi-agent framework harnessing [NVIDIA Agent Toolkit](#), including NVIDIA Nemotron, to route tasks to the best models and drive accurate, efficient and autonomous enterprise workflows.

Hitachi is developing physical AI technologies to address real-world operational challenges by using NVIDIA Nemotron and [NVIDIA Cosmos™ open models](#), along with its proprietary information technology (IT) and operational technology (OT) domain knowledge. As part of a multi-agent orchestration platform, these technologies are designed to connect and coordinate IT and OT operations, helping transform enterprise-scale business processes across complex workflows.

Sakana AI is collaborating with NVIDIA to integrate NVIDIA Nemotron into its Fugu model-orchestration platform, expanding the range of AI models Fugu can intelligently orchestrate to dynamically select the best model for each task in agentic AI workflows. By routing each request to the model best suited for the job, Fugu helps developers balance accuracy, performance and cost across multiple open and proprietary AI models. Fugu demonstrates how thoughtful orchestration can unlock capabilities beyond what an individual model achieves on its own. [Early performance results](#) on complex, real-world coding tasks reinforce the promise of coordination as a path to more capable AI.

Open and Customizable, Deployable Anywhere

Nemotron models are released with open weights, datasets and recipes, giving organizations the transparency and control to customize models for domain-specific workflows and deploy them where their applications and data reside.

Developers can use NVIDIA NeMo to customize, evaluate and optimize models for their use cases, and deploy them in environments that meet regulatory, sovereignty and data localization requirements.

Nemotron models are available on [Hugging Face](#), [ModelScope](#), [OpenRouter](#) and [build.nvidia.com](#) as NVIDIA NIM™ microservices, and through [NVIDIA Cloud Partners](#), inference platforms and cloud service providers.

About NVIDIA

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

Certain statements in this press release including, but not limited to, statements as to: Together with Japan's AI leaders, NVIDIA advancing an open AI ecosystem that accelerates discovery, strengthens national capability and ensures every society can participate in — and benefit from — the AI revolution; expectations with respect to growth, performance, availability, and benefits of NVIDIA's products, services and technologies, and related trends and drivers; expectations with respect to NVIDIA's third party arrangements, including with its collaborators and partners; expectations with respect to technology developments, and related trends and drivers; projected market growth and trends; expectations with respect to AI and related industries; and other statements that are not historical facts are forward-looking statements within the meaning of Section 27A of the Securities Act of 1933, as amended, and Section 21E of the Securities Exchange Act of 1934, as amended, which are subject to the "safe harbor" created by those sections based on management's beliefs and assumptions and on information currently available to management and 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 and political conditions; NVIDIA's reliance on third parties to manufacture, assemble, package and test NVIDIA's products; the impact of technological development and competition; development of new products and technologies or enhancements to NVIDIA's existing products and technologies; market acceptance of NVIDIA's products or NVIDIA's partners' products; design, manufacturing or software defects; changes in consumer preferences or demands; changes in industry standards and interfaces; unexpected loss of performance of NVIDIA's products or technologies when integrated into systems; NVIDIA's ability to realize the potential benefits of business investments or acquisitions; and changes in applicable laws and regulations, 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.

© 2026 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, Nemotron, NVIDIA Cosmos, NVIDIA HGX, NVIDIA Jetson, NVIDIA NeMo and NVIDIA NeMo Retriever 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.

Natalie Hereth

Corporate Communications
NVIDIA Corporation
press@nvidia.com