

ServiceNow and NVIDIA Fuel a New Class of Intelligent AI Agents Across the Enterprise

2025-05-06

New April Nemotron 15B reasoning model delivers lower latency, lower inference costs, and faster agentic AI—purpose built for performance, cost, and scale

ServiceNow brings accelerated data processing to Workflow Data Fabric with the integration of NVIDIA NeMo microservices, driving a closed-loop data flywheel process that enhances model accuracy and personalized user experiences

LAS VEGAS--(BUSINESS WIRE)-- **Knowledge 2025** – Today at ServiceNow’s annual customer and partner event, **Knowledge 2025**, **ServiceNow** and NVIDIA announced an expansion of their partnership to fuel a new class of intelligent AI agents across the enterprise. This includes the debut of a new high-performance ServiceNow reasoning model, April Nemotron 15B—developed in partnership with NVIDIA—that evaluates relationships, applies rules, and weighs goals to reach conclusions or make decisions. The open-source LLM is post-trained with NVIDIA and ServiceNow-provided data, helping deliver lower latency, lower inference costs, and faster agentic AI. The companies also unveiled plans to bring accelerated data processing to ServiceNow **Workflow Data Fabric** with the integration of select **NVIDIA NeMo microservices**, driving a closed-loop data flywheel process that enhances model accuracy and personalized user experiences.

The April Nemotron 15B reasoning model represents a significant step forward in developing compact, enterprise-grade LLMs purpose-built for real-time workflow execution. The model was trained using **NVIDIA NeMo**, the NVIDIA Llama Nemotron Post-Training Dataset, and ServiceNow domain-specific data with **NVIDIA DGX Cloud** on Amazon Web Services (AWS). It delivers advanced reasoning capabilities in a smaller size—making it faster, more efficient,

and cost-effective to run on NVIDIA GPU infrastructure as an **NVIDIA NIM** microservice. Benchmarks show promising results for the model's size category, reinforcing its potential to power agentic AI workflows at scale. The debut of this model comes as enterprise AI continues to rise as a transformative force—helping businesses address growing complexity, navigate macroeconomic uncertainty, and drive smarter, more resilient operations.

To support ongoing model innovation and AI agent performance, ServiceNow and NVIDIA also unveiled a new collaboration on a joint **data flywheel** architecture that will integrate ServiceNow Workflow Data Fabric and select NVIDIA NeMo microservices. This integrated approach curates and contextualizes enterprise workflow data to refine and optimize reasoning models, with guardrails in place to help ensure that customers are in control of how their data is used and processed in a secure and compliant manner. This enables a closed-loop learning process that improves model accuracy and adaptability—accelerating the development and deployment of highly personalized, context-aware AI agents designed to enhance enterprise productivity.

“With this new Apriel Nemotron 15B reasoning model, we’re powering intelligent AI agents that can make context-aware decisions, adapt to complex workflows, and deliver personalized outcomes at scale,” said Jon Sigler, EVP of Platform and AI at ServiceNow. “But the model is just one part of the innovation. Our collaboration building a data flywheel—powered by Workflow Data Fabric and NVIDIA NeMo—enables a virtuous cycle of learning and improvement. This helps us build AI agents that are contextually aware, deeply personalized, and aligned to the real-time needs of the enterprise.”

“NVIDIA and ServiceNow share a mission to reimagine employee productivity through AI tools that help people get more done,” said Kari Briski, Vice President of Generative AI Software for Enterprise at NVIDIA. “Together, we’ve built the Apriel Nemotron 15B model to serve as an enterprise-grade reasoning engine and plan to integrate NVIDIA NeMo microservices into ServiceNow Workflow Data Fabric, providing a powerful foundation for intelligent digital agents.”

The new Apriel Nemotron 15B reasoning model and data flywheel integration will better equip AI agents to meet the growing demands of customers with continuous data and process feedback. For example, imagine an AI agent resolving a complex billing issue by pulling in past customer interactions, reasoning through the problem, and recommending the next best step—getting faster, more accurate, and more efficient with every case it handles. Together, ServiceNow and NVIDIA are turning enterprise data into real-time, personalized action.

This latest milestone builds on the **recently announced** AI agent evaluation tools and integration of **NVIDIA Llama Nemotron** models with the ServiceNow AI Platform to accelerate agentic AI development. ServiceNow and NVIDIA have a shared vision of designing innovations that ensure LLMs—and the experiences they’re powering—are not only intelligent, but also measurable, secure, and ready for real-world deployment. The co-development of the Apriel Nemotron 15B reasoning model and data flywheel integration marks a natural next step in the companies’

deep partnership—furthering their collaboration to power enterprise workflows with greater speed, precision, and cost-efficiency.

Availability

The Apriel Nemotron 15B model is expected to be available in Q2 2025.

Additional information:

Learn more about ServiceNow **Knowledge 2025**.

About ServiceNow

ServiceNow (NYSE: NOW) is putting AI to work for people. We move with the pace of innovation to help customers transform organizations across every industry while upholding a trustworthy, human centered approach to deploying our products and services at scale. Our AI platform for business transformation connects people, processes, data, and devices to increase productivity and maximize business outcomes. For more information, visit: www.servicenow.com.

Forward-Looking Statements

This press release contains “forward-looking statements” about the expectations, beliefs, plans, and intentions relating to new innovations ServiceNow and NVIDIA are co-developing. Such statements include statements regarding future product capabilities and offerings and expected benefits to ServiceNow. Forward-looking statements are subject to known and unknown risks and uncertainties and are based on potentially inaccurate assumptions that could cause actual results to differ materially from those expected or implied by the forward-looking statements. If any such risks or uncertainties materialize or if any of the assumptions prove incorrect, ServiceNow’s results could differ materially from the results expressed or implied by the forward-looking statements made. ServiceNow undertakes no obligation, and does not intend, to update the forward-looking statements. Factors that may cause actual results to differ materially from those in any forward-looking statements include: (i) delays and unexpected difficulties and expenses in executing the product capabilities and offerings, (ii) changes in the regulatory landscape related to AI and (iii) uncertainty as to whether sales will justify the investments in the product capabilities and offerings. Further information on factors that could affect ServiceNow’s financial and other results is included in the filings ServiceNow makes with the Securities and Exchange Commission from time to time.

are trademarks and/or registered trademarks of ServiceNow, Inc. in the United States and/or other countries. Other company names, product names, and logos may be trademarks of the respective companies with which they are associated.

Media Contact

Katlyn Hirokawa

408-489-7381

press@servicenow.com

Source: ServiceNow