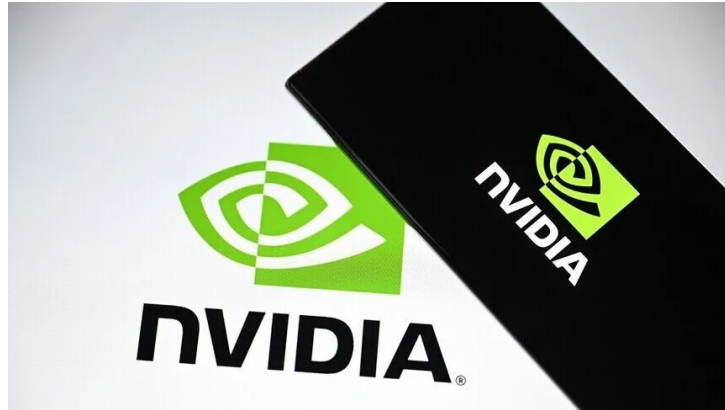


---

## Nvidia unveils lightweight AI model for autonomous agents



Published on August 12, 2026

Document Date: Tue, Aug 11 2026 10:49:41 pm

Category: ,English,International - ,Snippets

Show on website : [Click Here](#)

---

rki.news|Sources Anadolu

SAN FRANCISCO, Aug 11: Nvidia has introduced a lightweight open artificial intelligence model designed to support high-volume tasks performed by autonomous AI agents, while requiring significantly less computing power.

The company said its new Nemotron 3.5 Lightning model has 30 billion total parameters but activates only 3 billion for each task through a mixture-of-experts architecture.

The model can operate on a single supported GPU system, including Nvidia's DGX Spark and H100, and offers a context window of up to 1 million tokens, according to Nvidia's model documentation. It is designed for applications including code review, tool use, security-alert monitoring and customer billing queries within larger AI-agent systems.

Nvidia said its benchmark tests showed the model could generate output up to four times faster and complete agentic tasks 30% faster than other open models in its class.

The company released the model's weights under its OpenMDW 1.1 license, allowing developers and businesses to customize it with their own data and workflows.

Nvidia also announced NeMo Switchyard, an open-source routing library designed to direct AI tasks to suitable models based on capability and cost.

Separately, Nvidia said partnerships with six major financial institutions could mobilize more than \$500 billion in third-party capital for AI infrastructure over time.