NVIDIA Launches OpenShell Reference Architecture for Safe AI Agents

Platform Brings Together NVIDIA Vera Rubin, Vera CPUs and BlueField-4 to Power Enterprise Infrastructure to Build, Run, Observe and Enforce Safe Agents at Industrial Scale

News Summary:

- The NVIDIA OpenShell Reference Architecture delivers enterprise governance, isolation and security for building, containing, observing and enforcing safe agents across enterprise infrastructure and edge systems.
- Built on the foundation of NVIDIA OpenShell secure runtime software, the architecture brings together NVIDIA Vera Rubin infrastructure for model inference, NVIDIA Vera CPUs for agent harnesses and tool use, and NVIDIA BlueField-4 DPUs running NVIDIA Sentry for in-silicon enforcement. The OpenShell Reference Architecture extends to the edge with NVIDIA RTX Spark, NVIDIA DGX Spark and NVIDIA DGX Station systems.
- The OpenShell Reference Architecture runs on leading enterprise operating systems, including Canonical, Red Hat, and Microsoft Windows and Microsoft Execution Containers.
- Anthropic, AWS, Cisco, CrowdStrike, Dassault Systèmes, Dell Technologies, Google Cloud, HPE, IBM, Meta, Microsoft, OpenAI, Palo Alto Networks, Perplexity, SAP, ServiceNow, Siemens, SpaceX Super Intelligence and others are among those working with the OpenShell Reference Architecture to enhance the security of agent deployments.

GTC Berlin--NVIDIA today announced the NVIDIA OpenShell Reference Architecture, which enables building, containing, observing and enforcing safe agents on enterprise infrastructure and edge systems.

The OpenShell Reference Architecture builds on the foundation of NVIDIA OpenShell™ secure runtime software, extending it into full-stack, industrial-scale infrastructure for agent development and operations.

It includes NVIDIA Vera Rubin infrastructure for running enterprise workloads, including closed and open models built with NVIDIA Nemotron™; NVIDIA Vera CPUs for operating agent harnesses; open source NVIDIA OpenShell software; and NVIDIA BlueField® DPUs for running the NVIDIA Sentry unified security software framework that provides continuous out-of-band monitoring of agent behavior and in-silicon policy enforcement.

The OpenShell Reference Architecture is extensible to agents running at the edge across industries such as retail, healthcare, manufacturing, robotics, software development and telecommunications with NVIDIA RTX Spark™, NVIDIA DGX Spark™ and NVIDIA DGX Station™ systems.

“AI capability and safety must advance together," said Jensen Huang, founder and CEO of NVIDIA. “As enterprises prepare to deploy billions of AI agents, safety and control must be built into the computing architecture. NVIDIA created OpenShell to give agents the freedom to do useful work within clearly defined boundaries -- contained, monitored and under control. With the OpenShell Reference Architecture, agents can think on Rubin, act on Vera and be governed on BlueField."

Full-Stack Infrastructure Built for Agents
The OpenShell Reference Architecture features NVIDIA accelerated computing paired with software for secure agent operations.

- NVIDIA Vera Rubin, with NVIDIA Vera Rubin NVL72 infrastructure at its core, is fungible across pretraining, post-training and inference. For agentic workloads, NVIDIA Vera Rubin NVL72 systems deliver 30x higher throughput per megawatt.
- NVIDIA Vera CPUs deliver 2x the performance of x86 processors with greater energy efficiency and run NVIDIA OpenShell, which provides a secure runtime boundary for autonomous agents.
- NVIDIA BlueField-4 DPUs continuously monitor agent behavior with NVIDIA Sentry, built on NVIDIA DOCA™ software. BlueField and NVIDIA Sentry together provide in-silicon security enforcement -- if an agent attempts to move outside its software boundary, NVIDIA Sentry quarantines and stops it in milliseconds. NVIDIA Sentry, powered by DOCA and running on BlueField-4, boosts agent capacity by up to 20%.

NVIDIA Partners Building Secure Agent Deployments With OpenShell
NVIDIA is working with industry leaders building infrastructure, models and software to enable industries to safely deploy agents in production.

The OpenShell Reference Architecture runs on leading enterprise operating systems and container orchestration layers, including Linux and Kubernetes infrastructure powered by Canonical, SUSE and Red Hat, as well as Microsoft Windows using Microsoft Execution Containers (MXC). It can run on cloud operating systems from Amazon Web Services, Google Cloud and Microsoft Azure.

Enterprises can run agents from model builders including Anthropic, Cohere, Google, Meta, OpenAI, Perplexity, Reflection and SpaceX Super Intelligence on OpenShell infrastructure. NVIDIA is also working with model builder partners on OpenShell deployments for advanced model research.

Cybersecurity leaders 1Password, Check Point, CrowdStrike, Cisco, F5, Fortinet, Microsoft, Okta, Palo Alto Networks, TrendAI and Zscaler are integrating support for the OpenShell Reference Architecture into their agent security solutions.

Leading banks -- including Europe's BBVA and Santander -- are working with NVIDIA to explore the OpenShell Reference Architecture as a framework for strengthening agent safety, security and governance.

Enterprise software leaders Cadence, Dassault Systèmes, IBM, SAP, Siemens and Synopsys are working with OpenShell to build agent software platforms. Palantir and NVIDIA are working with customers to deploy secure agents in production.

SAP is integrating the NVIDIA OpenShell Reference Architecture with Joule Studio runtime, part of SAP Business AI Platform, connecting business governance with secure agent execution.

ServiceNow today announced the general availability of ServiceNow Arc, a personal long-running AI agent with actions governed at runtime by the ServiceNow Super Intelligence Control Tower and enforced by NVIDIA OpenShell. In addition, ServiceNow's recently announced AI Passport authority layer carries those policies to NVIDIA OpenShell and NVIDIA Sentry, so enterprises can revoke an AI agent's access across software and hardware within seconds.

NVIDIA system manufacturers Dell Technologies, Cisco, HPE, Lenovo and Supermicro are among those building infrastructure with the OpenShell Reference Architecture.

NVIDIA is also working with cloud service providers including CoreWeave, Crusoe, Firebird, Firmus, GMI Cloud, IREN, Lambda, Nebius, Nscale, Oracle Cloud Infrastructure and Vultr to deploy OpenShell-ready infrastructure.

Inference and post-training service providers including Baseten, Deepinfra, Fireworks and Together AI are building OpenShell Reference Architecture integrations into their model development pipelines.

Availability
The NVIDIA OpenShell Reference Architecture is now available.

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: 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, and related deployments, availability and timing; expectations with respect to technology developments, and related trends and drivers; expectations with respect to enterprise adoption and deployment of AI agents and related safety, governance and control; 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 NVIDIA'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.

Many of the products and features described herein remain in various stages and will be offered on a when-and-if-available basis. The statements above are not intended to be, and should not be interpreted as a commitment, promise, or legal obligation, and the development, release, and timing of any features or functionalities described for our products is subject to change and remains at the sole discretion of NVIDIA. NVIDIA will have no liability for failure to deliver or delay in the delivery of any of the products, features or functions set forth herein.

© 2026 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, and other NVIDIA product and service names are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries and regions. Other company,

product, and service names may be trademarks of the respective companies with which they are associated.

Corporate Communications
NVIDIA Corporation
press@nvidia.com