

s →

Hot Stocks →

48,977.181.23%▲

10 6,902.050.64%▲

Q23,395.820.69%▲

Fear & Greed Index

47

Lates

Mark

laid o

News

What

next f

tech i

made

world

most

valua

comp

Big O

doesi

share

Trum

drear

makin

Vene

oil gr

again

Sams

new p

looks

straig

of sci

fiction

got to

BUSINESS > TECH • 4 MIN READ

Nvidia just laid out what’s next for the tech that made it the world’s most valuable company

2 HR AGO
By Lisa Eadicicco



≡ **CNN Business**

Sign in



Nvidia founder and CEO Jensen Huang speaks at CES 2026 in Las Vegas on January 5, 2026. ...



Las Vegas — Nvidia just provided a closer look at its new computing platform for AI data centers, Vera Rubin, a release that could have major ramifications for the future of AI given the industry's massive reliance on the company's tech.

Nvidia previously announced some details about Vera Rubin but laid out how the system will work and revealed its launch timing during the CES tech conference in Las Vegas on Monday. Vera Rubin is currently in production and the first products running on it will arrive in the second half of 2026, the company said.

Nvidia has become the poster child for the AI boom, with the pervasiveness of its AI chips and platforms propelling it to briefly become the world's first \$5 trillion company last year. But the company is also combatting **fears of an AI bubble** amid growing competition and a push by tech companies to make their own AI chips to decrease reliance on Nvidia.

Advertisement

Nvidia CEO Jensen Huang, clad in his signature leather jacket, addressed the question of where AI funding is coming from – a point central to the bubble debate – in his opening remarks on stage at the theater inside the Fontainebleau Las Vegas. He said companies are shifting budgets in research and development in classical computing methods to artificial intelligence.

“People ask, where is the money coming from? That’s where the money is coming from,” he said.

The Vera Rubin platform is an attempt by Nvidia to position itself as the answer to the computing challenges posed by increasingly demanding AI models – such as whether **existing infrastructure** can handle increasingly complicated AI queries. The company claims in a press release that its upcoming AI server rack, called Vera Rubin NVL72, “provides more bandwidth than the entire internet.”

With Vera Rubin, Nvidia says it’s developed a new type of storage system to help AI models process more complex, context-heavy requests more quickly and capably. Existing types of storage and memory used by traditional computers and even the graphics processing units powering data centers won’t be enough as companies like Google, OpenAI and Anthropic shift from offering simple chatbots to full-fledged AI helpers.

Advertisement

Huang walked through what the transition from chatbots to agents on Monday. In a video demonstration, a person built a personal assistant by connecting a friendly-looking tabletop robot to multiple AI models running on Nvidia's DGX Spark desktop computer. The robot was able to do things like recount the user's to-do list and even tell the dog to get off the couch.

Huang said creating such an assistant would have been unimaginable several years ago but is "utterly trivial" now that developers can rely on large language models rather than traditional programming tools to build apps and services.

In other words, the old way simply won't cut it as AI grows more sophisticated and "reasons" on tasks that take multiple steps like these, Nvidia claims.

"The bottleneck is shifting from compute to context management," Dion Harris, Nvidia's senior director of high-performance computing and AI hyperscale solutions, said on a call with reporters ahead of the press conference.

Advertisement

"Storage can no longer be an afterthought," he added.

Nvidia also just entered into a licensing agreement with a company called **Grog** that specializes in inference ahead of CES, another sign that it's investing heavily in that branch of AI.

"Instead of a one-shot answer, inference is now a thinking process," Huang said, referring to the process AI models go through to "think" and "reason" through answers and accomplish tasks.

All of the major cloud providers – Microsoft, Amazon Web Services, Google Cloud and CoreWeave – will be among the first to deploy Vera Rubin, Nvidia said in its press release.

Computing companies like Dell and Cisco are expected to incorporate the new chips into their data centers, and AI labs such as OpenAI, Anthropic, Meta and xAI are likely to embrace the new tech for training and to provide more sophisticated answers to queries.

Advertisement

Nvidia also **deepened its push** into autonomous vehicles with new models called Alpamayo and “physical AI,” the type of AI that powers robots and other real-world machinery, building on the vision it laid out during its GTC conference in October.

But Nvidia’s progress and prevalence also means it shoulders the burden of consistently surpassing Wall Street’s high expectations and assuaging concerns that spending on AI infrastructure is far outpacing tangible demand.

Meta, Microsoft and Amazon, among others, have spent **tens of billions in capital expenditures** this year alone, and **McKinsey & Company** expects companies to invest nearly \$7 trillion in data center infrastructure globally by 2030. And much of the support being poured into AI seemingly involves a relatively small group of companies trading money and technology back and forth in what’s known as “circular funding.”

The tangled web of AI investment

One expert estimated that every dollar spent on artificial intelligence generates \$8 of investment between the companies. Explore some of the biggest deals.

Click a company to highlight its deals

Note: Not an exhaustive list of AI company investments or deals.

Source: Securities and Exchange Commission; OpenAI; Amazon; Coreweave; Oracle; Nvidia; CNN reporting.
Graphic: Alex Leeds Matthews, CNN

Google and **OpenAI** have also been leaning into developing their own chips, allowing them to better tailor hardware to the specific needs of their models. Nvidia has also been facing **growing competition from AMD**, and chipmaker Qualcomm also recently announced it's getting into the data center business.

“Nobody wants to be beholden to Nvidia,” Ben Barringer, global head of technology research at investment firm Quilter Cheviot, said in a previous CNN interview when asked about other companies like **Google potentially challenging Nvidia in AI chips**. “They are trying to diversify their chip footprint.”



Up next



How AI shook the world in 2025 and what comes next

6 MIN READ



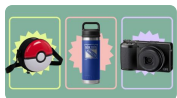
We tested more than 100 tech products in 2025. Here are our 20 favorite

2 MIN READ



The Microsoft Surface Pro 2024 is one of my favorite tablets ever, no thanks to AI

15 MIN READ



Our favorite product releases this week: Igloo, Ricoh, Yeti and more

8 MIN READ



Our favorite product releases this week: Athleta, Black+Decker, Philips Hue and more


10 MIN READ



I like the Google Pixel 10 Pro, but it might not be the Pixel to buy this year

17 MIN READ

Most popular

- 1** Pentagon moves to cut Sen. Mark Kelly's military retirement pay as punishment over 'illegal orders' video
- 2** Trump said Venezuela stole America's oil. Here's what really happened
- 3** Minnesota Gov. Tim Walz ends reelection bid amid welfare fraud scandal
- 4** Cuba is left reeling in the aftermath of the Venezuela strike
- 5** The surprising US plan in Venezuela comes with huge risks for Trump
- 6** One person in custody after incident at Vance's Ohio residence
- 7** 'Grab what you can:' The global rush for second passports
- 8** Hear what happened in court as Maduro pleads not guilty  **3:57**

9 From Greenland to Iran: Trump's threats stretch far and wide since his Venezuela strike

10 Minnesota investigators say child care centers accused of fraud are operating normally as governor drops reelection bid

Search CNN...

Sign in

Live TV

Listen

Watch

World

US Politics

Business

Markets

Health

Entertainment

Tech

Style

Travel

Sports

Science

Climate

Weather

Ukraine-Russia War

Israel-Hamas War

Features

Watch

Listen

Games

About CNN

Business

FOLLOW CNN BUSINESS



DOWNLOAD THE CNN APP



Most stock quote data provided by BATS. US market indices are shown in real time, except for the S&P 500 which is refreshed every two minutes. All times are ET. Factset: FactSet Research Systems Inc. All rights reserved. Chicago Mercantile: Certain market data is the

property of Chicago Mercantile Exchange Inc. and its licensors. All rights reserved. Dow Jones: The Dow Jones branded indices are proprietary to and are calculated, distributed and marketed by DJI Opco, a subsidiary of S&P Dow Jones Indices LLC and have been licensed for use to S&P Opco, LLC and CNN. Standard & Poor's and S&P are registered trademarks of Standard & Poor's Financial Services LLC and Dow Jones is a registered trademark of Dow Jones Trademark Holdings LLC. All content of the Dow Jones branded indices Copyright S&P Dow Jones Indices LLC and/or its affiliates. Fair value provided by IndexArb.com. Market holidays and trading hours provided by Copp Clark Limited.

[Terms of Use](#) [Privacy Policy](#) [Manage Cookies](#) [Ad Choices](#) [Accessibility & CC](#) [About](#) [Newsletters](#) [Transcripts](#)

© 2026 Cable News Network. A Warner Bros. Discovery Company. All Rights Reserved.

CNN Sans [™] & © 2016 Cable News Network.

