

AI

DeepSeek Sparked a Market Panic. Here Are the Facts.

By [Tae Kim](#) [Follow](#)

Jan 27, 2025 4:29 pm EST

In this article

- ↓ COMP
- ↑ NVDA
- ↑ META
- ↑ TSLA
- ↑ MSFT



*** ONE-TIME USE *** Traders and financial professionals work on the floor of the New York Stock Exchange (NYSE) at the opening bell on January 27, 2025 in New York City. (ANGELA WEISS/AFP/GETTY IMAGES)

Social media never lets facts get in the way of a good story.

Over the weekend, viral posts [suggested](#) that Chinese company DeepSeek had recreated OpenAI's artificial intelligence prowess for just \$6 million, versus the billions spent by U.S. tech giants. The runaway hype quickly raised questions about America's AI leadership and tanked tech stocks on Monday. The [Nasdaq Composite](#) ↓ **COMP -3.07%** finished the day down 3.1%, while AI leader [Nvidia](#) ↓ **NVDA -16.97%** tumbled 17%.

But the reality is far more complicated. DeepSeek didn't simply replicate OpenAI's ability by spending a few million dollars.

DeepSeek first [unveiled](#) the \$6 million figure in a late December technical paper for its DeepSeek-V3 model. The start-up estimated the model's final training run, taking 2.8 million GPU hours, would cost \$5.6 million if it rented that amount of cloud capacity. Importantly, DeepSeek excluded costs related to "prior research and ablation experiments on architectures, algorithms, or data."

This means the number omits all R&D funds spent developing the model's architecture, algorithms, data acquisition, employee salaries, buying GPUs, and test runs. Comparing a theoretical final run training cost with overall U.S. company spending on AI infrastructure capital expenditures is comparing apples and oranges. DeepSeek's overall cost is likely much higher.

On Monday, Bernstein analyst Stacy Rasgon cited DeepSeek's disclosure, noting a "fundamental misunderstanding" over the \$5 million figure. It is "categorically false that China duplicated OpenAI for \$5 million."

Technology fund manager Gavin Baker [called](#) using the \$6 million training figure "deeply misleading," emphasizing that a smart team

couldn't train the DeepSeek model from scratch with a few million dollars.

Several AI experts strongly suspect that DeepSeek used advanced U.S. model outputs in addition to its own to optimize its models through a process called distillation, improving smaller models' capability by using larger models.

Recent news out of China, meanwhile, debunks the idea of AI on the cheap. Last week, China [announced](#) plans to provide \$137 billion in financial support for AI over the next few years. DeepSeek founder Liang Wenfeng reportedly [told](#) Chinese Premier Li Qiang last week that American export restrictions on AI GPUs remained a "bottleneck," according to The Wall Street Journal.

This all means that global technology companies are likely to keep spending on AI infrastructure to train new advanced models and develop the next generation technology.

Amid the DeepSeek frenzy, [Meta Platforms](#) ↑ META +1.91% CEO Mark Zuckerberg [announced](#) on Friday his company would invest \$60 billion to 65 billion on capital expenditures this year while significantly growing its AI teams. Last October, Meta gave guidance for 2024 capex of \$38 billion to \$40 billion. "This will be a defining year for AI," Zuckerberg wrote Friday on Facebook, adding that Meta is building a 2+ gigawatt data center and will have over 1.3 million GPUs by year-end.

To be clear, there are important things to be gleaned from DeepSeek. In the wake of its new models, there are new questions about computing capacity needed for AI inference, the process of generating results from AI models.

The Chinese start-up innovated by using techniques like “Mixture of Experts,” making its smaller distilled models extremely efficient at inferencing. New Street Research says using DeepSeek-V3 for inference costs about 90% less than comparable OpenAI models.

Nvidia shares fell Monday on fears and uncertainty over future demand for training and inference. An Nvidia spokesperson called DeepSeek “an excellent AI advancement and a perfect example of Test Time Scaling,” noting “inference requires significant numbers of Nvidia GPUs and high-performance networking.”

DeepSeek could mean that AI models are becoming vastly more efficient. If so, U.S. vendors are likely to ultimately replicate that work. But that doesn’t automatically mean a sudden and permanent glut of AI chips. History suggests that tech innovation fills in available supply.

“Deep Learning has a legendary ravenous appetite for compute, like no other algorithm ever developed in AI,” [said](#) [Andrej Karpathy](#), [Tesla’s](#) [↓ TSLA -2.32%](#) former AI director and an OpenAI co-founder. “I would never bet against compute as the upper bound for achievable intelligence in the long run.”

Former Intel CEO Pat Gelsinger [offered a similar view](#). “Computing obeys the gas law. Making it dramatically cheaper will expand the market for it. The markets are getting it wrong, this will make AI much more broadly deployed.”

Meanwhile, [Microsoft](#) [↓ MSFT -2.14%](#) CEO Satya Nadella [referenced](#) the Jevons paradox, an 1865 observation by an economist that efficiency improvements increase consumption as additional use cases become

economical and are discovered. “As AI gets more efficient and accessible, we will see its use skyrocket, turning it into a commodity we just can’t get enough of,” he said.

If AI becomes more capable, developers and enterprises will eventually find ways to use it, short-term bumps aside. This time isn’t likely to be different.

Write to Tae Kim at tae.kim@barrons.com 