

AI Just Got 90 to 99.99 Percent Cheaper

Stargate for \$500 billion or DeepSeek for \$5 million?

GALE POOLEY
JAN 29, 2025

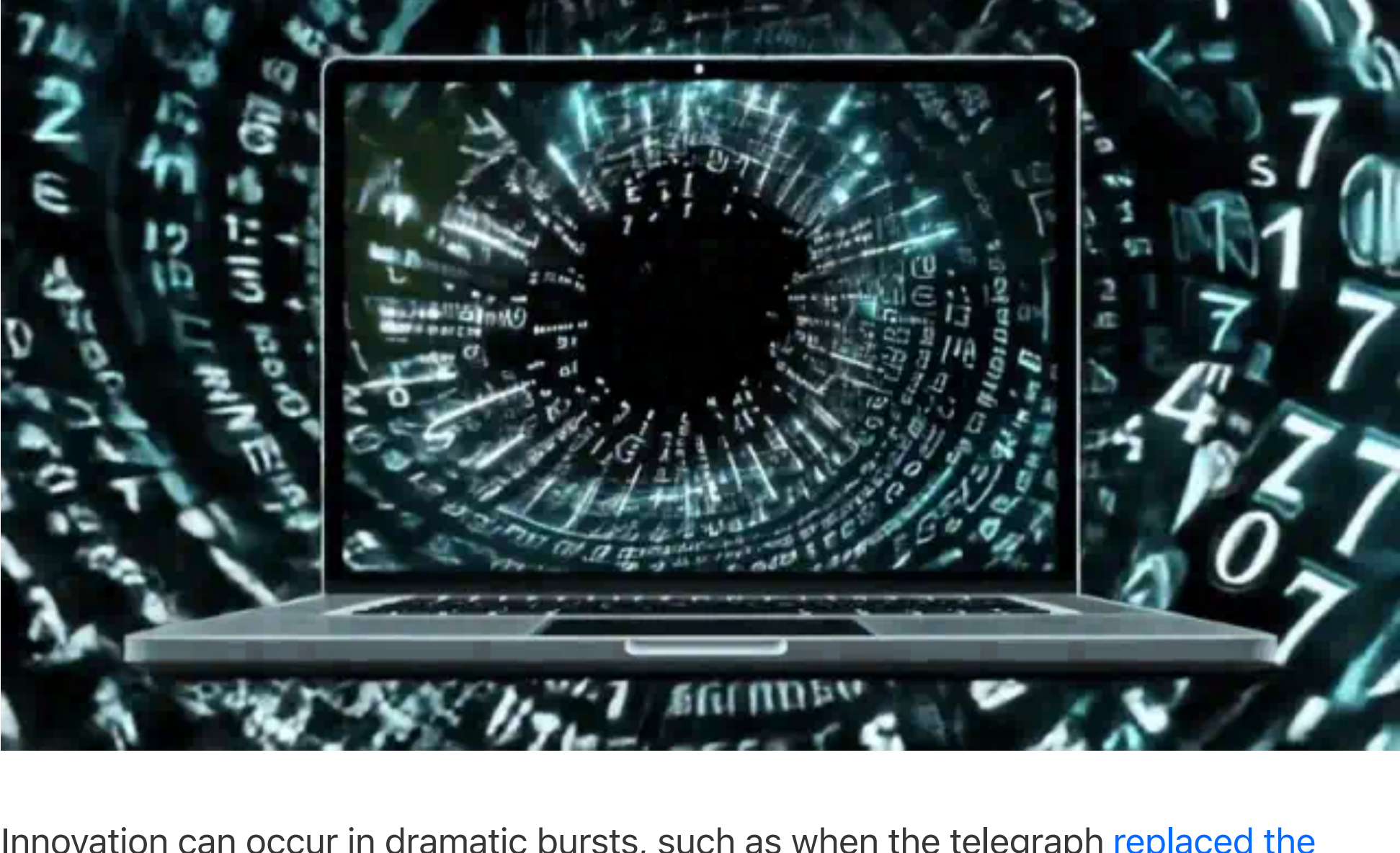
♡ 6

💬 2

🔄 1

Share

⋮



Innovation can occur in dramatic bursts, such as when the telegraph [replaced the Pony Express](#). This iconic mail carrier cut previous delivery times in half and reigned for 18 months as the fastest way to deliver information across the United States. The Pony Express was introduced on April 3, 1860, and delivered mail between St. Joseph, Missouri, and Sacramento, California. The 2,000-mile route took approximately 10 days, with riders traveling 75 to 100 miles each and switching horses every 10 to 15 miles.

Western Union erected the first telegraph poles on July 4, 1861. Construction took 112 days to complete the first electronic transcontinental communication system on Oct. 24, 1861. Two days later, the Pony Express was discontinued.

The telegraph reduced the time it took to deliver a message by 99.93 percent, from 10 days to 10 minutes. Sending a message got 1,439,900 percent faster.

Are we witnessing another ponies to electrons innovation between OpenAI and DeepSeek? Maybe.

Thanks for reading Doomslayer! [Subscribe for free](#) to receive new posts in your inbox.

✓ [Subscribed](#)

[Peter Diamandis](#) has noted that DeepSeek was founded less than two years ago. With only 200 employees and \$5 million, DeepSeek has developed a new artificial intelligence (AI) system. By comparison, OpenAI was founded 10 years ago, has around 4,500 employees, and has raised \$6.6 billion in capital. AI tech giants like OpenAI and Anthropic have been spending \$100 million or more to train their models. DeepSeek has matched their systems for 95 percent of the cost. A 95 percent drop in cost means you now get 20 for the price of one, indicating a 1,900 percent increase in abundance. DeepSeek has done this with three innovations:

1. Precision reimagined. Instead of using computational overkill (32 decimal places), they proved that 8 decimal places is enough. The result? 75 percent less memory needed. Sometimes the most powerful innovations come from questioning the most basic assumptions.

2. The speed revolution. Traditional AI reads like a first-grader: “The . . . cat . . . sat . . .” But DeepSeek’s multitoken system processes whole phrases simultaneously: 2 times faster with 90 percent accuracy. When you’re processing billions of words, this is transformative.

3. The expert system. Instead of one massive AI trying to know everything (imagine one person being a doctor, a lawyer, *and* an engineer), they built a system of specialists. Traditional models rely on 1.8 trillion parameters being active *all the time*. DeepSeek, by contrast, relies on 671 billion in total, but only 37 billion are active at once (97.9 percent fewer).

Diamandis goes on to note more staggering results from DeepSeek’s innovations:

- training costs slashed from \$100 million to \$5 million;
- GPU requirements slashed from 100,000 GPUs to 2,000 GPUs;
- 95 percent reduction in API costs;
- runs on gaming GPUs instead of specialized hardware; and
- done with a team of less than 200 people, not thousands.

The DeepSeek system is open source, which means anyone can verify, build upon, and implement these innovations. You can download the new app on [your iPhone](#).

Bonus: AI now has a counterpoint to the environmentalists who say AI uses so much electricity. DeepSeek just brought down the cost of inference by 97 percent.

Cathie Wood at ARK Investment [has observed](#) “over the last few years that AI training and inference costs have been dropping 75% and 85-90%, respectively. DeepSeek may be accelerating the pace of change, but the declines already are dramatic. Faster cost declines will add to demand, more for inference, a more competitive GPU space than training, and one of the nuances.”

Moore’s law suggests that computer transistor abundance doubles every two years. That would indicate a compound rate of around **41.4** percent a year. The cost to [train](#) an AI system to recognize images fell 99.59 percent from \$1,112.64 in 2017 to \$4.59 in 2021. This would indicate a compound rate of **295** percent a year. AI is growing over seven times faster than Moore’s law.

Nvidia is leading the development of these systems and their CEO Jensen Huang has [claimed](#) that AI processing performance has increased by “no less than one million in the last 10 years.” This is a compound annual rate of 298 percent. He expects this rate to continue for the next 10 years. That would mean we go from one to one trillion in twenty years. We’ll be 976 million times ahead of Moore’s law. Quite astonishing.

The Pony Express needed lots of fast horses and skinny riders. The telegraph was a whole new platform that used wire and batteries and poles instead. DeepSeek may be the Western Union to the Pony Express.

So what [about Stargate](#)—a \$500 billion US AI infrastructure initiative led by OpenAI’s Sam Altman, Oracle’s Larry Ellison, and Softbank’s Masayoshi Son? They want to spend 100,000 times more than DeepSeek has spent so far. Will their product be 100,000 times more valuable?

Microsoft’s CEO [Satya Nadella](#) brought up Jevon’s paradox in regard to DeepSeek. On January 26, 2025, on X, he said: “Jevons paradox strikes again! 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.”

On January 6, [Nvidia announced](#) its new Nano line of AI development hardware starting at \$259 and its new Project DIGITS as the “world’s smallest AI supercomputer capable of running 200B-parameter models” and expected to be priced at around \$3,000. Between DeepSeek’s open-source software and Nvidia’s hardware, the world could experience a brilliant efflorescence of superabundance in learning.

We expect to see AI make dramatic advances in our ability to discover valuable new knowledge, but we’ll also come to realize that the only intelligence in artificial intelligence so far has been human intelligence. If human beings have the freedom to innovate, the potential to create resources is infinite.

Thanks for reading Doomslayer! This post is public so feel free to share it.

Share