# THOUGHTS ON THE DEEPSEEK SHOCK TO EQUITY MARKETS

Last week, a relatively unknown Chinese firm released a remarkably advanced open-source artificial intelligence (AI) model called DeepSeek R1. They claim to have trained it for less than $6 million, which is a fraction of the $100 million to $1 billion reputed costs of training state-of-the art US models.

By many benchmark measures, DeepSeek R1 appears to provide similar performance to OpenAI's latest highly capable o1 model, although it still lags behind their most advanced (but not yet released) o3 model (**Figure 1**). The news that DeepSeek has rapidly become the top-downloaded free app on Apple's App Store caused a sharp drop in tech stocks by raising questions about the sustainability of current AI-related investment trends.

BY WILLIAM P. STERLING, PH.D.

Global Strategist

FIGURE 1

DeepSeek Announcement On X.com
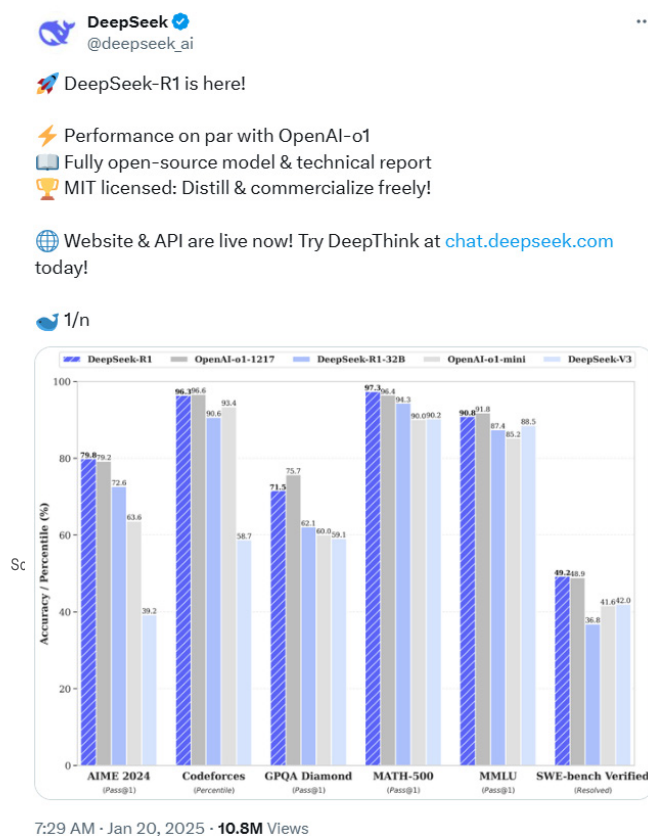
Source: X.com, January 20, 2025.

By open sourcing their model, much as Meta does with their own AI models, DeepSeek may prove to be a formidable threat to the revenue models of the leading US AI providers since "free" is a compelling price point for state-of-the-art performance.

As legendary Silicon Valley investor Marc Andreessen put it: "DeepSeek R1 is one of the most amazing and impressive breakthroughs I have ever seen — and as open source, a profound gift to the world."

Some observers suspect that they actually spent far more money than $6 million on pre-training research and development and may have had access to as many as 50,000 Nvidia H100 GPUs in developing their product.[1] We may never know the truth on that issue, or whether they had ample resources from China's government to help them do an end run around US export controls on Nvidia's most advanced GPUs, which are critical to AI research and development.

What is clear is that DeepSeek's software engineering expertise appears to be world class and that, despite US export controls, China may be only a few months behind the leading US AI providers in providing cutting-edge AI capabilities. In addition, much of their progress appears to be based on impressive innovation in algorithms, spurred by their relative lack of access to the most advanced AI hardware.

It should be no surprise that algorithmic advances are capable of delivering order-of-magnitude gains in the so-called "effective compute" that AI systems are based on. Indeed, Nvidia's Jensen Huang has referred many times to why AI could become a million times more powerful over the next decade due to dual progress in both hardware and software.[2]

However, the shock has been how rapid China's progress has been and whether the efficiency gains on the algorithm side mean that investors will have to substantially mark down their projections for capital spending on the "pick and shovel" AI providers like Nvidia and AMD or on related power providers involved in the coming data-center buildout.

WE SUSPECT SUCH FEARS ARE OVERBLOWN FOR A NUMBER OF REASONS

1. We think the $6 million development cost figure is highly misleading and that there were other substantial costs involved, as noted above.

2. We think the genuine advances represented by DeepSeek's engineering are a proof statement that AI is rapidly advancing to so-called AGI (artificial general intelligence) or even ASI (artificial super intelligence), which would have transformative growth impacts for the global economy.

3. By suggesting more rapid AI progress in the pipeline the race to scale up AI systems could intensify rather than slow down, since the biggest prizes will be for those who develop AGI or ASI first.

4. Indeed, Microsoft's CEO, Satya Nadella, hinted at this logic by bringing up the so-called "Jevons paradox," noting that improved efficiency in AI may boost its demand. And while this could eventually make it a commodity, it could also lead to increased energy consumption as AI performs a wider range of tasks.[3]

5. The recently announced Stargate Project by OpenAI, Oracle, Softbank, and MGX, with partners like Microsoft and Nvidia, still seems likely to go ahead with estimated outlays of $500 billion over the next four years as the race to scale continues. If anything, the news that China is so close to the US in terms of AI capabilities raises the stakes for the US to achieve AGI first. It's also worth noting that Microsoft just reiterated its commitment to $80 billion of capital spending while META very recently increased its 2025 capital expenditures to $60 – $65 billion, up from $38 – $40 billion in 2024. The DeepSeek news likely came as no surprise to the US AI leaders as DeepSeek V3 was released in late December and is likely incorporated into the latest thinking on capital spending plans of US hyperscalers.

When faced with volatile markets, it is important to keep economic fundamentals in mind. We recently asked leading AI researcher Tamay Besiroglu how he saw AI investment spending evolving over the next few years.[4] His response is still worth keeping in mind:

---

[1] Scale AI CEO Alexandr Wang told Andrew Ross Sorkin of CNBC's *Squawk Box* on January 24, 2025, that "DeepSeek has about 50,000 Nvidia H100s that they can't talk about because of the US export controls that are in place." If so, at $35,000 per H100 GPU, that would amount to access of $1.75 billion in compute power.

[2] Hasan Chowdhury, "Nvidia boss Jensen Huang predicts AI power will increase a 'millionfold' in a decade," *Business Insider*, November 19, 2024.

[3] Satya Nadella, "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." Posted on X.com, January 27, 2025.

[4] William Sterling, "What's Next for AI: Interview with AI Researcher Tamay Besiroglu." GW&K Global Perspectives, October 2024.

"There's a possibility that spending on compute-related capital — including semiconductors, data centers, and dedicated energy infrastructure — will significantly exceed current estimates. Industry leaders like Sam Altman and Satya Nadella have suggested that scaling current AI techniques by approximately 1,000 times could potentially automate a substantial portion of human labor. If this perspective proves accurate, it might justify investments far beyond even the rumored $100 billion data centers."

"One intuition of this is just to consider that the global wage bill currently stands at approximately $60 trillion per year. Thus, developing an artificial intelligence system capable of capturing even 10% of this labor value over a few years could potentially yield returns in the tens of trillions of dollars. Admittedly, this would require widespread automation, given that capturing 10% of the value of labor will require the automation of much more than 10% of tasks humans can do. However, this suggests if there is a good chance of developing AI systems that can flexibly substitute for human labor, investing in compute-related capital on the scale of a trillion dollars could potentially offer a favorable return on investment."

In short, we believe that the race to AGI is still on, and the US remains in the lead. The latest news from China will almost certainly intensify the competitive dynamics. As is always the case in competitive markets, that leaves open the question of who will be the ultimate winners and losers. So while this news does raise competitive risks for some of the producers of AI models and infrastructure, it also raises the potential for most other companies and consumers to be big beneficiaries from cheaper AI models becoming embedded in products, services, and software.

William P. Sterling

William P. Sterling, Ph.D.
Global Strategist