

DeepSeek: Cutting Through the Noise & Key Takeaways

Omega Venture Partners

February 2025

DEEPSEEK: CUTTING THROUGH THE NOISE & KEY TAKEAWAYS

DeepSeek has been at the center of AI headlines this past week—some claims accurate, others exaggerated. Here's our expert perspective as to its broader implications for the AI landscape, what matters, and why the emergence of DeepSeek further validates Omega's investment strategy and focus.

Reinforcing Omega's Investment Focus

DeepSeek's emergence reshapes AI market dynamics to some extent, but the biggest opportunity is not in training large models. The real value lies in AI applications—leveraging these new, cheaper, and more powerful models to drive real-world impact. As the AI stack matures, capital-intensive model training is becoming commoditized, while the highest returns will come from companies that apply AI in transformative, industry-specific ways.

DeepSeek's Breakthrough: Efficiency Without Tradeoffs?

DeepSeek's models work very well and demonstrate impressive inference efficiency, running at significantly lower cost than top OpenAI models. This challenges two widely held beliefs:

- **Bigger foundation models aren't necessarily better**—DeepSeek shows that smaller, optimized architectures can match or exceed performance in certain tasks.
- **Reasoning doesn't have to be expensive**—Unlike traditional AI models, DeepSeek's inference is designed to be more cost-effective, without sacrificing quality.

What remains unverified? While inference efficiency is demonstrably strong, DeepSeek's claimed training costs lack transparency. The reported \$5.7M cost to train its model is difficult to reconcile with what we know about compute requirements, making it likely that other expenses and full hardware stack data were not accurately disclosed.

DeepSeek's Threat to Proprietary Foundation Models

This is bad news for OpenAI. OpenAI has justified its premium pricing in foundation models by positioning its 'o1' offerings as a generation ahead, largely due to their use of test-time compute—where models engage in additional reasoning during inference, such as when generating responses, to improve output quality.

DeepSeek erodes OpenAI's pricing power by delivering comparable inference efficiency at a fraction of the cost—undermining one of OpenAI's key competitive advantages.

Hardware & Infrastructure Impact: Unlikely to Decline

Some argue that DeepSeek's efficiency and claims will drastically reduce the need for GPUs, but our view is different—we believe Jevons Paradox applies: as AI becomes cheaper, more efficient, and more accessible, overall demand is likely to increase, not decrease.

- **Cheaper AI = More AI.** As AI models become more efficient and cost-effective, more companies will build and train their own AI, increasing overall compute demand.
- **Specialized AI will flourish.** Instead of relying on a few general-purpose models, expect more domain- and industry-specific AI solutions.
- **NVIDIA & Compute Infrastructure Players** are likely going to be fine. As AI becomes cheaper to build and deploy, more businesses will enter the race, expanding GPU demand rather than reducing it.

This efficiency-driven expansion follows Jevons Paradox—an economic principle stating that as efficiency increases, costs decrease, and overall demand rises rather than declines ([link](#)).

Scrutinizing DeepSeek's Claims

While DeepSeek inference efficiency is impressive, DeepSeek's claims about the cost and resources used in training its models should be taken with a huge grain of salt.

- **The \$5.7M training cost—misleadingly incomplete.** DeepSeek has reported the cost of a single training run, but this excludes critical expenses such as infrastructure (CapEx for GPUs), human labor, and the cost of multiple iterations required to refine the model and architecture. The true cost of developing DeepSeek's model is almost certainly far higher.
- **Only H800 GPUs? Unclear.** Given existing sanctions & geopolitical factors, it's unlikely to expect them to disclose their real or full hardware stack.
- **Potential Data Exfiltration from US Models? Unverified.** OpenAI has an incentive to push this claim, but no clear evidence exists yet. If true, it would dramatically reduce DeepSeek's costs (and violate OpenAI's terms of service).

AI Progress Is Accelerating, Not Slowing

2024's AI plateau media narrative was wrong. DeepSeek proves that smarter architectures continue unlocking major efficiency gains—cost, compute, and reasoning are all improving in parallel. DeepSeek's architecture and algorithms demonstrate **several key ideas**:

- **Mixture of Experts (MoE)** – Rather than relying on a monolithic model, DeepSeek distributes tasks among multiple specialized "expert" models to improve efficiency. While MoE has been explored before (e.g., Databricks and Mistral), prior implementations struggled to achieve state-of-the-art performance. DeepSeek's reported success is intriguing but unproven—it remains to be seen whether they have meaningfully advanced the approach or if current results will hold at scale.
- **Multi-Head Latent Attention (MLA)** – AI models use attention mechanisms to process different parts of an input sequence. DeepSeek's key innovation is in the 'latent' component—designed to significantly reduce memory overhead, making the model more efficient without sacrificing performance. If this technique proves effective, it could enable smaller, faster models without compromising comprehension.
- **Multi-Token Prediction (MTP)** – Allows the model to predict multiple words at once, dramatically increasing response speed.
- **The 'Think' Phase with Reinforcement Learning** – Dynamically applies additional reasoning during inference only when needed, optimizing computational efficiency. You can see this in action by testing DeepSeek ([link](#)).

As we've long predicted, foundation models are evolving toward greater efficiency. DeepSeek's architectural choices—when integrated with other test-time compute techniques—could unlock even greater reasoning efficiency and cost reductions across the entire ecosystem of foundation models.

The bigger picture: AI innovation is not just about throwing more computing power at the problem—better architecture matters just as much, if not more.

AI Innovation Knows No Borders

There's been much discussion about DeepSeek in the context of US-China AI competition. The AI research community isn't focused on national rivalries—it is optimizing for cost, efficiency, and performance in foundation models.

The idea that governments can control AI breakthroughs is naïve — I think we've intuitively known that all along, but this just underscores that:

- Foundation model progress is driven by open innovation—not regulatory constraints.
- Government containment is insufficient to prevent foundation model innovation.

This also underscores why open-source foundation models are structurally advantaged. DeepSeek, like other major breakthroughs, leverages open research and contributes back—a challenge to proprietary, closed-model approaches like OpenAI and Anthropic.

Positive Tailwind for AI Adoption

We believe AI-powered applications will see an acceleration in deployment as lower model costs remove barriers to adoption. As AI becomes more efficient, it will also be far more widely adopted.

- **More companies will build AI-powered products** at lower costs.
- **More specialized AI models** will emerge, unlocking **new industry-specific applications**.
- **The AI compute ecosystem is set for continued expansion.**

Omega's Investment Strategy—Focusing on Backing Companies That Harness AI to Drive Real Business Value at Scale

The emergence of DeepSeek further validates Omega's investment strategy—focusing on backing companies that harness AI to drive real business value at scale. By investing in AI applications with clear economic impact, Omega is focused on backing the companies best positioned to become the consequential, enduring winners of tomorrow.

DISCLOSURES

These materials and the views expressed herein are intended for informational purposes only, and should not be used as or as a substitute for investment, financial, tax or other forms of advice or for any other purpose. These materials are not intended to act as an offer or solicitation for the purchase or sale of any financial instrument. Information referenced or used herein has been obtained from sources believed to be reliable but OMEGA does not warrant the completeness or accuracy of any third-party information. Past performance is not indicative of future results. Graphs and charts cannot be used in and of themselves to determine the advisability of investing in securities or the timing of the purchase or sale of any security.

These materials contain opinions and/or estimates of OMEGA and its personnel that constitute its and their perspectives as of the date of these materials and are subject to change without notice. The opinions contained in these materials do not take into account individual circumstances, objectives, or needs and are not intended as recommendations of particular securities, financial instruments, or strategies. The recipient of these materials must independently undertake its own diligence and thereafter make its own decisions regarding the advisability of any securities or financial instruments referenced herein.

OMEGA will not be responsible or have any liability for any injuries or damages of the recipient of these materials or any third party, including injuries or damages that may result from (i) the reliance of any person upon any information or opinions provided herein or (ii) any errors, inaccuracies, omissions in, or any other failure of, the data herein, from whatever cause. OMEGA shall not be liable for direct or incidental special or consequential damages resulting from the information contained herein and/or from any decision taken on the basis of such information, regardless of whether such damages were foreseeable. Some of the companies selected for inclusion in this report are held by investment vehicles or clients managed by OMEGA.



<https://www.omegavp.com>



<http://linkedin.com/company/omegavp>



<https://x.com/omegavp>