Deepseek and its Impact on Power Demand

On January 20, 2025, Chinese AI startup DeepSeek released DeepSeek-R1, an open-source reasoning model apparently matching performance of leading US AI models like OpenAIâ??s o1. What makes this development remarkable is not only that it emerged from China â?? a country previously perceived as lagging in the AI race â?? but also that it was allegedly developed at a cost of just USD 5.6 million, a fraction of the expenditure required for US models. This is particularly notable given it was allegedly developed using inferior chips as the sale of Nvidiaâ??s most advanced chips to China has been banned since October 2023.

Much has been written about DeepSeek and its potential implications for global technology companies, as well as for firms supplying equipment and power to data centers. However, it is important to clarify a key point: the USD 5.6 million figure represents only the cost of releasing this final iteration of the model and does not account for the substantial capital invested in earlier stages of development. To draw an analogy, it is like claiming that the only effort required to run a marathon is the effort expended during the final 42 kilometers, ignoring the months or years of training that precede the race.

Additionally, DeepSeek employs a technique called distillation, which transfers knowledge from larger pre-trained models developed by companies like OpenAI, Meta, and Anthropic. Without these foundational models and their enormous development costs, DeepSeek-R1 could not have been trained for USD 5.6 million. There has also been speculation that DeepSeek may have received support from the Chinese government, something which might explain that the release date coincided with Trumpâ??s inauguration.

That said, the true cost of DeepSeek-R1 is somewhat beside the point. What matters is that the world now has access to a powerful open-source model that can be distilled onto smaller models at minimal expense. Moreover, DeepSeek appears to have made significant advancements in computational efficiency, representing a step change in the field. Improved efficiency implies lower power requirements, which is why the grid sector and other AI enablers experienced a sharp sell-off in January.

Was this sell-off justified? We believe so. The key question now is whether power demand estimates for AI data centers will be revised downward. On this, we are less certain. Initial updated forecasts suggest that power demand growth remains robust.

We expect DeepSeekâ??s advancements to spur the development of more purposeful AI applications, likely increasing demand for inference data centers relative to those focused on large language model (LLM) training. However, it remains unclear whether this shift will reduce or increase overall demand for compute power.

One telling indicator is the capex budgets of hyperscalers, some of which were announced after DeepSeekâ??s release, have increased significantly beyond market expectations, with no indication that less money will be spent for the same level of compute power. If anything, the opposite seems true, at least in the near term.

The longer-term outlook is more uncertain. We believe DeepSeek-R1 marks the beginning, rather than the end, of efficiency-driven breakthroughs in the race toward Artificial General Intelligence (AGI). This is undoubtedly positive news for AI adopters â?? companies leveraging AI technologies. Cheaper AI applications will democratize access, accelerate adoption, and likely boost global productivity.

The bigger question is what this means for the enablers, the companies supplying chips, infrastructure, and power. Will Microsoft CEO Satya Nadellaâ??s recent post on X prove prescient: â??Jevons paradox strikes again!â?•? This 160-year-old economic principle, which posits that improvements in resource efficiency lead to increased rather than decreased resource consumption, has seen a resurgence in discussions over the past few weeks. In the context of AI, cheaper and more efficient models could lead to even greater investment and adoption, driving up demand for resources like power and infrastructure.

We believe this dynamic will hold true over time, though it is difficult to predict whether adoption will keep pace with efficiency improvements on a month-to-month or year-to-year basis.

In light of these developments, we believe the â??powering AIâ?? trade â?? a popular theme and part of the fundâ??s exposure â?? deserves a gentle reset. The valuation multiples at which some of these companies traded were difficult to justify based solely on grid and renewable energy demand. While the anticipated growth in AI data centers was expected to turbocharge demand for electrical components and related services, the increased uncertainty warrants a more cautious approach to valuations.

Having sharply reduced our net exposure to this thematic, we believe a reset will create more attractive entry points for the fund to rebuild positions in the future.

Date Created

2025/08/26