Deeper thinking around DeepSeek

By Victor Chung, Julia Frayer, and Donald Osborne Moss at London Economics International LLC (www.londoneconomics.com)

Ten days ago, a new AI model with a sinister name emerged and sent shockwaves through the tech world, and by extension, set back expectations for the growth of the US electricity sector.

What do we know about this new AI model?

- Built in China by the startup DeepSeek, the new, open source AI model DeepSeek R1 is a lot smaller than the large language models developed in the United States (such as OpenAI's GPT4-o1, which powers the ChatGPT app). The largest DeepSeek-R1 model involves 685 billion parameters, much smaller than GPT4-o1 with an estimated1.8 trillion parameters.

- Preliminary feedback from users suggests that DeepSeek R1 performs similarly to the leading US-made large language models, in terms of reasoning abilities and AI-assisted services.

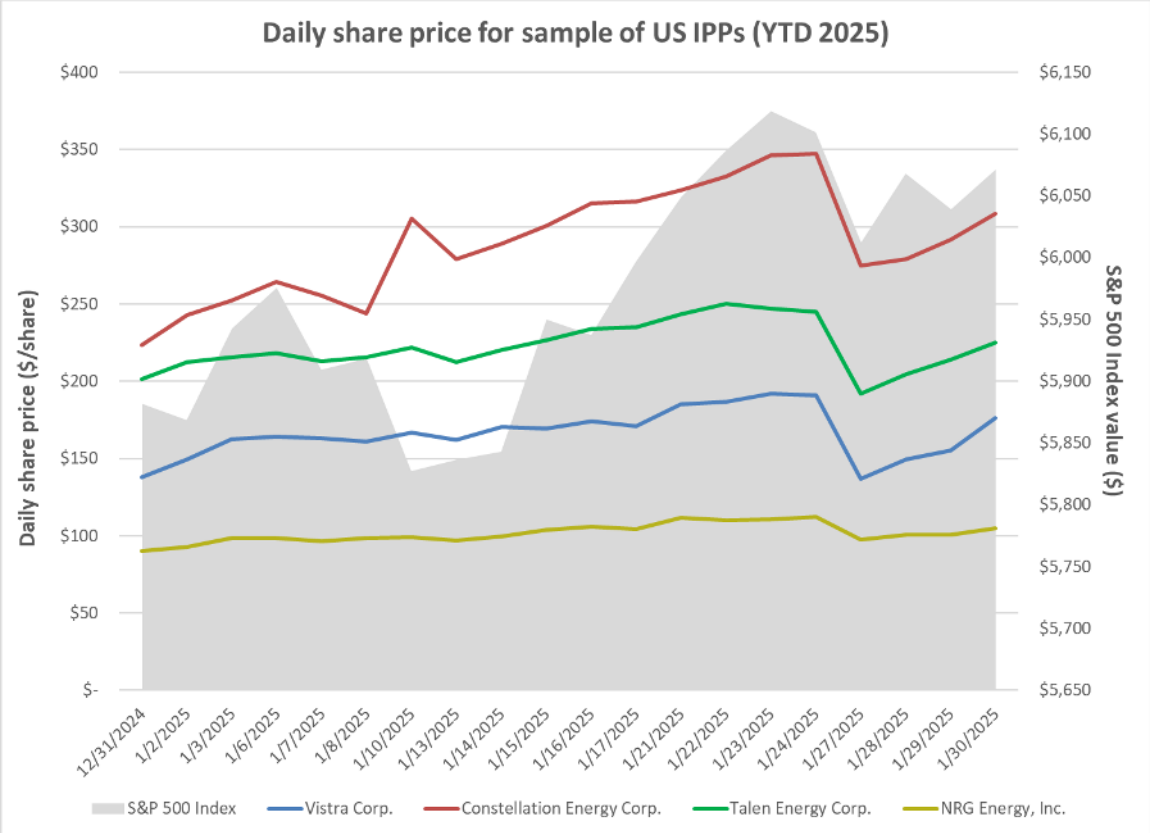
- DeepSeek represented the model's training cost at under $6 million. In contrast, OpenAI reportedly spent $80 million to train the GPT-4 model, while Google's Gemini Ultra was reported to cost over $190 million to train. This means the initial fixed costs for launching DeepSeek R1 would have been a fraction of what the US companies spent on the development of similarly performing models.

What is the connection to the electricity sector?

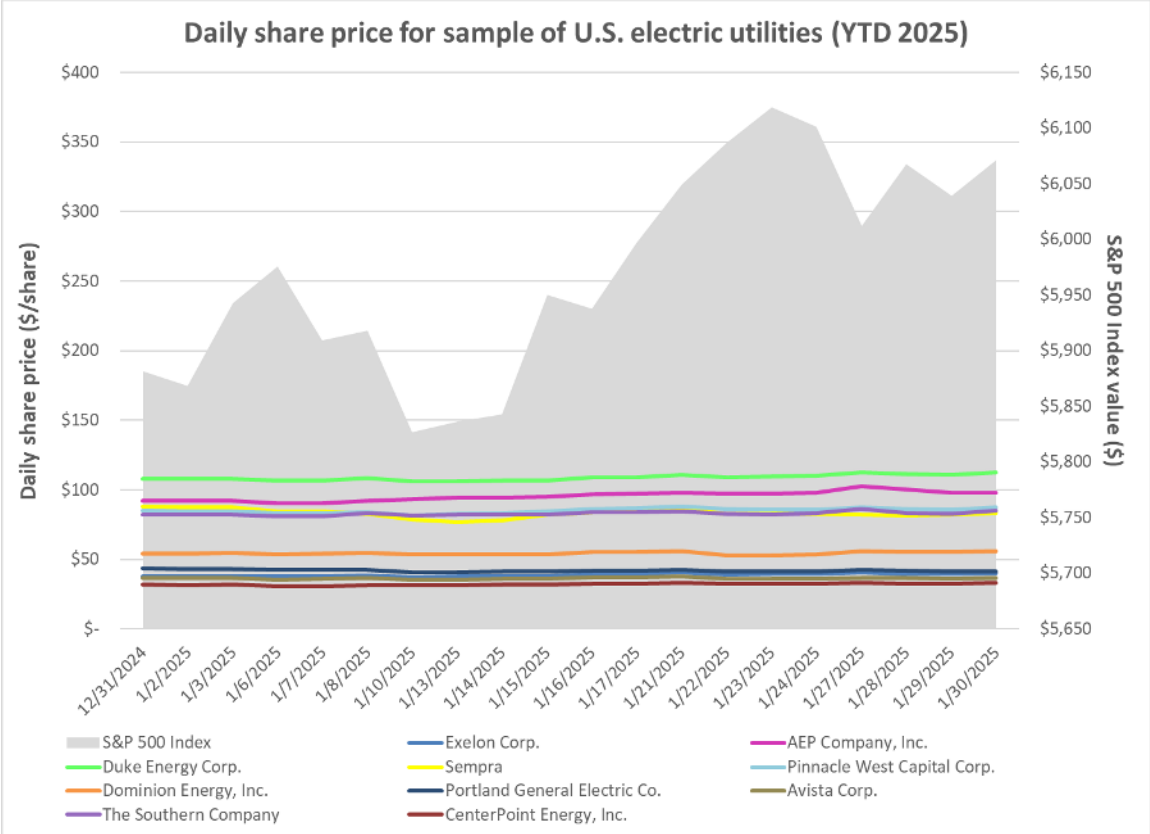
Training large language models requires computers, and computers need electricity. The lower training cost (and shorter time frame for development) of DeepSeek R1 means it used less electricity, as compared to other large language models. The improvement in energy efficiency surrounding the release of DeepSeek R1 raises important questions about how much electricity would be needed to support an AI-driven economy in the medium and longer term.

The US electric industry was expecting significant incremental electricity demand before the release of DeepSeek R1. Based on data collected by London Economics International LLC, Regional Transmission Organizations (" RTOs") and independent system operators ("ISOs") across the US projected over 65 GW of incremental electricity demand due to data center growth by 2030 (this estimate is on the low end as it excludes regions of the US, such as New England and New York, that have not (yet) seen material levels of data center specific customer interconnection requests). The Lawrence Berkeley National Laboratory recently estimated that by 2028, total power demand for data centers will range between 72 and 132 GW. Overall, such levels of additional electricity demand could not be met with existing infrastructure - significant investments in electric power generation and transmission and distribution delivery infrastructure would be needed to reliably meet such levels of incremental demand.

DeepSeek's speed of development and associated energy efficiency should cause us to question these electricity demand forecasts. And the financial markets did just that. Stock prices of independent power producers took the brunt of the market reaction to the news, especially in key areas of the United States where data center growth expectations were a major driver of future electricity demand. Share prices of Constellation Energy (CEG), Vistra Corp (VST) and Talen Energy (TLN) dropped over 20% between January 26th and January 27th, while NRG Energy (NRG) dropped by 15%, as seen in the chart below, wiping out most of the stock price appreciation these companies enjoyed from the prior year. The markets had positioned these generation-owning companies as the primary beneficiaries of higher electricity demand given the location of their power generation fleets and business activities to attract data center business to their respective markets.



Interestingly, the stock prices of major investor-owned utilities with electric transmission and distribution networks in these same key areas were not as adversely affected (as seen in chart below), despite of the fact that these utilities were also expected to have to make longer term grid investments to support new data center customers flocking to these regions.



What does this recent experience imply for forecasting growth in US electricity demand ?

1. **The electricity sector will be increasingly sensitive to developments in the AI sector.** Over the last 30 years, the electricity demand across the United States has grown at less than 1% per annum. In December 2024, the North American Reliability Council ("NERC") published its annual 10-year reliability assessment and projected an overall electric load growth expectation that is more than 50% higher than the historical average, primarily driven by the growth in data centers. The technical developments associated with the speed and cost of training DeepSeek's large language model raise important questions about the accuracy of previously published long term electricity demand forecasts.

2. **Electricity use during model development is only half the equation.** Large language models continue to use electricity even after they are "trained." Every time a user asks AI-powered app to solve a problem or do a task, the underlying large language model must make an "inference." That takes computing power and uses electricity. Model size has generally been viewed as a major factor contributing to measured energy use during inference. DeepSeek-R1, which has only 685 billion parameters, is significantly smaller than other leading large language models. However, independent research has also pointed to the importance of model architecture in determining energy use during the inference process. As more analysis becomes available about the quality of inferences from smaller models like DeepSeek R1, refinement of the electricity demand forecasts will be necessary

– that refinement needs to account for the electricity needs of the AI industry at both the development stage (when the model is trained) and deployment stage (when the model is asked to make repeated inferences).

3. **Technology improvements may lead to wider adoption with cascading effects on electricity demand.** Energy efficiency improvements translate into a lower cost of developing large language models. This will likely drive down the cost of AI-related services, making them accessible to more customers and spurring additional use cases in the longer term. Economic theory posits that wider adoption of AI-related services will increase the demand for data centers and drive up electricity demand. The magnitude of elasticity of demand for AI usage may be a novel consideration, but it is important that we start to explore how these technology effects factor into the size of the future AI economy and the demand for electricity over the long term. New electricity infrastructure takes time to build – investors, utilities, and regulators require electricity demand forecasts that are as accurate as possible.