DeepSeek and its Implications for the Future of AI Investing

REGIONS ASSET MANAGEMENT

February 2025

A special topic led by Fran Smitherman, Anjali Tagare and Chris Walker

On January 27, 2025, the rapid spread of news that a Chinese startup named DeepSeek had launched an AI assistant at a very low cost led to a dramatic sell-off in the shares of companies with ties to the AI growth story. Hundreds of billions of dollars in market capitalization were knocked out on the basis of a "shoot first, ask questions later" mentality that now, a few weeks later, the market and our team have had time to absorb and research more fully. Many of the stocks have recovered, but questions still remain about the dominance of AI as a driver of large capitalization US stocks over the foreseeable future. With this paper, we hope to dispel some of the fears and get some questions answered.

Why AI and the DeepSeek News Matter to RIM Investors

The RIM Equity research team has been exploring the topic of Artificial Intelligence for many years, and we have written several client-facing white papers on the topic.* In addition to that, we have exposure to the theme in many of the strategies in the Information Technology, Industrials, Communication Services and Utilities sectors. On the day of the DeepSeek selloff we saw the shares of companies such as Nvidia (NVDA) and GE Vernova down as much as 20% as investors reacted to a possible dent in their growth outlook which until that point, had been unquestioned.[1]

Since that day, most of the stocks that were impacted have recovered, with shares of Meta and Vistra up 20% year to date, GE Vernova up 13% year to date, and Quanta, Microsoft (MSFT), Google (GOOG) and Nvidia only down marginally year to date. As more questions arose regarding the claims that DeepSeek was making, the more investors' fears were allayed. Nevertheless, if hyperscaler spending were to fall over the next decade relative to expectations, the implications for the Technology, Industrials and Utilities companies will be worth following and being aware of.[2]

Artificial Intelligence cannot be ignored and while important investing themes must be explored, portfolios should also be diversified.

DeepSeek AI implications

DeepSeek R1, an AI reasoning model, was released on January 20, 2025. By January 27, 2025, it was the most downloaded app from the Apple (APPL) app store. This resulted in a panic sell-off of AI related companies.

1. ***Why the panic:*** DeepSeek creators implied that its V3 model was created in less than 2 months and cost less than $6MM, using older NVDA GPUs. This implies that advanced AI models can be built using less advanced/expensive AI chips. Also, DeepSeek R1 model is open source and therefore free for anyone to use. [3]
2. ***DeepSeek characteristics:*** DeepSeek R1 is the latest reasoning model which was built based on its DeepSeek V3 model. The V3 model is the large language model (LLM) and is similar to OpenAI's GPT-4, a general purpose model. DeepSeek claims to have trained its V3 model on about 2000 NVDA H800 GPUs for a cost of less than $6M in about 2 months. H800 is a modified NVDA H100 chip to address export controls. This is significantly lower than the approximately $63MM needed to train ChatGpt-4. [3]
3. ***Advantages of DeepSeek:*** By using software techniques like Mixture of Experts (MoE) and MLA (Multi-head Latent Attention), DeepSeek has shown that it is possible to train models without using expensive GPUs. It used simple reinforcement learning (RL) to train R1 instead of the traditional AI models which use human feedback (RLHF). It has also shown that knowledge distillation can be used to transfer knowledge from larger AI model, V3 in this case, to smaller model like R1. [3]
4. ***Disadvantages of DeepSeek:*** Most experts believe that the training costs for V3 are actually significantly higher than $6MM, which is the cost of the last training run and does not include any prior R&D costs. There is also some talk about usage of the latest NVDA chips to develop the training model. OpenAI claims

Authors: Fran Smitherman, CFA, Anjali Tagare, PH.D., Microbiology, & Chris Walker, CFA
Regions Investment Management, Director of Equity Reseach & Equity Analysts

that DeepSeek may have taken its data to develop its models. Another major disadvantage appears to be in training. While simple RL is cheaper than RLHF, there are several limitations. Simple RL can optimize training for specific tasks, its accuracy will be low for other tasks. RLHF training is done to closely follow human preferences. Therefore, it is best suited for AI models developed for human interactions. [3]

Investment Implications: There are two stages (and associated costs) to developing AI models. The first is to develop a large language model (LLM) which is referred to as the training model. The second is the inference model. The development of a training model takes a lot of compute and time. Also, high speed advanced GPUs are essential for developing new training models. By making inference training cheaper (in terms of compute) and faster, industry experts believe that this will lead to a push for edge AI, where a significant amount of computing is done on smaller devices like PCs and smart phones.

1. ***Semiconductors:*** Despite DeepSeek, most hyperscalers have increased their AI datacenter capex budgets for 2025. META, Amazon (AMZN), GOOG and MSFT expect to spend a total of $325.2B on AI infrastructure, a 46% increase over the $222.7B spent in 2024[4]. The U.S. government along with Oracle (ORCL), OpenAI and Soft Bank announced Project Stargate, a $500B[5] investment over the next four years, an initiative to advance AI infrastructure. Most analysts believe that cutting edge GPUs will continue to dominate training models. NVDA is the clear leader in GPUs for training models and is expected to maintain its leadership over the next couple of years. NVDA is also the leader in inference training. However, DeepSeek has shown a cheaper way for inference training. Thus, NVDA is expected to face competition from companies like AMD which make lower end GPUs.
2. ***Software:*** By lowering costs of inference training, software developers should be able to lower costs of integrating AI capabilities in their existing products. This should benefit software companies like MSFT, Salesforce (CRM), Adobe (ADBE) enabling them to lower costs for developing their specialized models. Lower costs will also accelerate AI adoption.
3. ***Hardware:*** By reducing compute requirements, development of edge AI is expected to accelerate. This should increase demand for AI enabled smartphones as well as PCs. AAPL has been a laggard in its AI capabilities. AAPL can now run powerful AI models on its hardware using its own chips which are optimized for high performance and energy efficiency.

Deepseek's Impact on Electrical Load Growth

A big part of the AI revolution within our investing thesis over the last year has been the expectation that utilities will be spending more money to expand the grid and to increase load growth for the data centers that are needed for the AI boom. However, US electrical load growth is much more than just data centers and in recent earnings calls, the Utilities companies and the Communication Services Companies have reiterated and/or increased their capex spending plans for the near term.

The International Energy Agency estimates from Sep'24 see data centers accounting for <5% of global electricity demand growth 2023-30. BofA forecasts US electrical load CAGR of 2.8% 2023-30, with data centers accounting for 50bps of incremental growth vs the 40bps historical CAGR seen during 2023-2023. BofA estimates gas turbine demand of 80GW+ 2025-2030; data centers accounts for ~10-15% of that, but ex-data center demand is still 2x the 10yr average. These figures are represented in Table 1.

Table 1:

Historical CAGR (2013-23)	**0.4%**
Data Centers	0.5%
Industrial growth	0.4%
EV adoption	0.5%
Building electrification	1.0%
Forecast CAGR (2023-30)	**2.8%**

Source: BofA Global Research, US Energy Information Administration

Data Center Infrastructure Benef ts from non-AI Demand

Cloud service providers continue to experience steady growth in non-AI revenue, with Microsoft's Azure non-AI revenue growth up 22% Y/Y as of the end of 9/30/24. Colocation firms make up ~50% of global data centers, and the North American colocation sites have just a 3% vacancy rate and where asking rents were 7% in 2024 according to CBRE. Industry insights suggest that colocation providers are facing capacity constraints due to prolonged lead times for both power and electrical equipment. From 2017 to 2022, U.S. data center construction spending grew at a 22% CAGR, even before the surge in AI driven demand.[6]

More compute remains our base case

Our core assumption is that AI-related computing and consequently power consumption, will continue to rise. Companies will keep reevaluating their capital expenditure allocations. Jon Gray, COO of Blackstone (owner of $80bn of data center assets) noted on January 30th: "Maybe there's a little less training that's done as a result of the less intensity. But at the same time, there's more inference, maybe there's more cloud, maybe there's more to do with enterprise. As usage goes up significantly, there's still a vital need for data centers, the form of that use may change… power usage… will continue because our lives are migrating online…there'll be even more questions coming even if the amount of power used on an individual question goes down."[7]

Hyperscalers are guiding to another year of strong capex growth

Recently, these companies reported earnings and projected 2025 capex of $335 billion, reflecting a ~45% increase, compared to a +50% increase in 2024. The growth outlook is strong across all six core companies, with increases ranging from +20-70%.[8]

Both near- and long-term data center project pipelines indicate a sustained buildout. In the U.S., $35B in data center investments are set to begin in 2025, up 24% from the $28B in 2024. Looking further ahead, the U.S. project pipeline totals $330B ($770B globally), reflecting ~90% year-over-year growth. For perspective, at the current construction spending rate of $30 billion per year, the U.S. pipeline extends through 2035 or roughly ~10 years.[9]

Since DeepSeek shook the market several weeks ago, electrical equipment stocks are down 15%. It took hyperscalers roughly 8 Years to see returns on SaaS/Cloud investments, and given the long-term nature of AI spending, the AI capex cycle is likely to continue through the end of the decade. As Nena Palmer, former Head of Data Center Portfolio Strategy at Google, explains: "Infrastructure is a long-range investment. If we're investing today, we may not be able to start to serve some of this compute, for, let's say 2 to 8 years. So I think a good benchmark here is to consider how long it took hyperscalers to become profitable in the cloud business. Across the board, that took around 8 years. So again, we're not going to start to realize, ROI until a number of years out."[10]

Conclusion

While the impact of inexpensive models such as DeepSeek could have far-reaching implications on the spending needs of the companies under our coverage, in the near term we think that the spending plans are locked in and necessary. There are questions regarding the legality of how DeepSeek did what it did (did they steal ChatGPT's intellectual property to train their models?) and there are also questions as to the actual cost they are claiming and whether it can be replicated. There is even thought that a more efficient use of a resource will lead to increased consumption, a positive, (Jevon's Paradox) as it will result in faster adoption.

We will be watching all of this closely over the months and year to come and continuing to invest in a way that we think is the most likely to outperform the markets as a whole in a risk-aware manner.

***Links to previous AI-adjacent papers written by the RIM Equity Research Team:**

- Edge AI – The Next Step in AI Infrastructure by Anjali Tagare -July 2024
- Revolutionizing Data Centers: A.I.'s Influence on Power Consumption by Chris Walker – February 2024
- Generative Artificial Intelligence: The Next Disruptive Technology – June 2023
-

Sources:

1. Factset Data
2. Factset Data
3. CNBC
4. https://finance.yahoo.com/news/chart-of-the-week-2025s-monster-ai-spend-is-now-revealed-110012820.html
5. https://www.forbes.com/sites/moorinsights/2025/01/30/the-stargate-project-trump-touts-500-billion-bid-for-ai-dominance/
6. Bank of America
7. Blackstone.com
8. Bernstein Research
9. Bernstein Research
10. Factset Data