Investment Insight

DEEPSEEK: NEW AI APPS AND THE INEXORABLE DEMAND FOR COMPUTING CAPACITY

We have held Nvidia in the J. Stern & Co. World Stars Global Equity Fund since early 2022. Given the market sell-off in AI-related stocks in late January we want to share our thoughts on AI, Nvidia and DeepSeek.

We remain very positive for Nvidia and its prospects for value generation. We believe that applications and use cases for AI have run far beyond available computing capacity and there is an extraordinary demand for computing capacity due to the rapid and growing adoption of AI by all industries. Over the past two years we have written several investment insights on AI – about the rise and significance of ChatGPT, why Nvidia is best placed to capitalise on the demand for AI, and why we continue to hold Nvidia despite its strong share price performance.

Nvidia has a very strong competitive position and an attractive business model with strong free cash flow generation. It is at the forefront of addressing the demand for computing capacity and we expect it to continue to deliver strong growth in revenues, profits and cash flows. Nvidia is only one of many of the companies we hold that will benefit from AI and we believe that the fundamental drivers for AI are unchanged.

On 27th January 2025, the AI exposed parts of the market sold off sharply, not just the technology hardware, semiconductor and networking companies, but also power and infrastructure companies. Some of these companies had share price falls of up to -30%. We think that the share price moves were exacerbated by passive ETF index investments and leveraged holdings. The decline was caused by the news that the Chinese app, DeepSeek, had topped the Apple download charts and so entered the mainstream news cycle.

Whilst the news has raised concerns, we believe that this kind of volatility is a normal part of investing in technology and innovation. If there is a further significant sell-off, then we believe that it will provide a very good buying opportunity.

DeepSeek

DeepSeek is a large language model (LLM) that has emerged out of a Chinese quantitative hedge fund and has released three products over the last month. The R1, the reasoning engine, was released on 20th January 2025 (the same day as President Trump's inauguration) and it has used a distillation technique that leverages other AI models such as OpenAI's ChatGPT and Meta's Llama3. The distillation technique leveraging other work is questionable and has been compared to reverse engineering, but it is building upon the progress that other LLMs have made.

DeepSeek has reportedly developed its LLM using ~2,000 GPUs which is significantly less than competing LLMs. For example, xAI is believed to have 100,000 GPUs and Meta up to 600,000. Some industry experts have made counter claims that DeepSeek are using many more GPUs than they are disclosing, perhaps up to ~50,000, and so it is difficult to verify what the actual figure is.

The DeepSeek R1 has performed well in tests, on an equivalent or better performance than the more recognised LLMs. It appears that DeepSeek may be more suited to technical questions like maths-based problems than language-based problems.

The improvements that DeepSeek has made include a technique called a 'Mixture of Experts,' where the problem is broken down with only specific parts of the model in operation at one time. This frees up capacity and results in far fewer GPUs being used to develop the model. It has also lowered the accuracy required in the model compared to the larger LLM models.

Based on the limited information available it cannot be verified if DeepSeek can compute as cheaply as claimed and if its performance is superior to other models. It is also unclear if it has commercial applications given concerns about applications provided by Chinese-based companies, including both B2B applications and B2C applications like TikTok.

Positive implications

Moreover, even if DeepSeek is as cheap and good as claimed, we think that it could be a positive not a negative for the technology sector and AI.

First, we have to remember that this kind of innovation is a normal part of technology development. One of the founding principles of technology investing is Moore's Law, which observes that transistor count per microchip doubles every two years and the cost per transistor decreases. This results in exponential growth in power and efficiency. Ultimately, technology is all about becoming more efficient and so we should not be surprised that alternative, more efficient AI LLMs have been created.

Secondly, it is also not the first time that a Chinese company has made a product much cheaper than American competitors (not just in technology).

Lastly, we think that Jevons Paradox is also applicable, where increased efficiency of a resource can result in higher overall consumption. This is most important for AI as the kind of simplification and efficiency promised by DeepSeek encourages greater use of AI and so even more compute consumption.

Nvidia in a short statement on 27th January, noted that the DeepSeek R1 is built upon time-test scaling, which involves greater compute processing at the inference stage. This means that the model re-prompts itself to think through the problem into a series of smaller ones, and so it may take longer to produce the answer and thus use more compute at the inference stage. The fact that DeepSeek appears to have used the distillation technique and leveraged off the other LLMs (it appears that in response to questions about its development, DeepSeek has replied that it is a model developed by OpenAI), means that is may have lowered the cost of the training phase, but that it might consume more compute in the inference phase. We have repeatedly said that for Nvidia, the next stage of AI demand will from the inference stage rather than the training model building phase. Therefore, we believe that the kind of development used by DeepSeek may not cause a decline in the demand for Nvidia's chips but actually an increase.

We have previously reduced the position of Nvidia during 2024, most recently in December at ~$135 in the World Stars UCITS fund. We have been proactively managing the position and lowered the exposure based on our assessment of portfolio risk and opportunities for reallocating capital to other positions.

However, we would also note that not all technology companies on the 27th January. Whilst the MSCI tech sector fell -4.7% and the semiconductor sector fell -7.7%, some technology companies rose, including our holdings in Salesforce +4% and Meta +2%. These companies will be beneficiaries of expanded AI. Salesforce in particular, has a significant proprietary data set that it can harness to deliver appropriate tailored AI solutions for clients. We think that it is possible that LLMs may end up as a commodity (an area that we are not invested in) and that the applications that sit on top of them deliver the real value of AI.

In summary, the technology sector is a fast-moving sector that requires active management. We note that just during the last week, Project Stargate was announced, an investment of up to USD 500 billion over four years to building new AI datacentres; Meta announced its intention to spend USD 60-65 billion on capex in 2025; and the Bank of China announced an allocation of 1 trillion Yuan to the AI industry over the next five years. These are all positives for AI development. We do not think that these projects would have been announced without any knowledge of DeepSeek and the innovation it claims to have achieved. The progress of AI remains intact and if there is any further significant share price sell off, it will provide a buying opportunity for the technology industry leaders.

Implications for Industrial holdings

In terms of the other exposures in the World Stars portfolio, the significant growth in data centre investments has led to an increased structural demand for grid infrastructure equipment, power management solutions, as well as high speed connectors, benefiting some of our Industrial holdings, most notably Eaton and Amphenol. This was evidenced for example in the 76% growth in demand within Amphenol's IT & Datacom business posted during the recently published 4th quarter earnings results. Similarly, in its FY2024 results, Eaton reported 45% revenue growth in Data Centres, including 75% with Hyperscalers.

We believe that investments into data centres will continue going forward for both traditional and AI-application specific data infrastructure. We have noted, however, that the valuation levels for both stocks had become elevated versus historical levels and our own absolute price targets leading us to reduce our position in both over the last year consistent with our investment approach.

Importantly both Eaton and Amphenol benefit from significant diversification in their end markets, with demand being driven by the transition of the global economy to net zero, the reshoring of critical industries, the digitalisation of industrial assets, increased defence spending globally, and robust commercial travel trends. This broad diversification across end markets is what underpins both companies' revenue trajectory for the years ahead independent of developments in any one end market.

Giles Tulloch and Katerina Kosmopoulou, CFA
28th January 2025