Global Artificial Intelligence: Thoughts on DeepSeek and potential implications

DeepSeek is an artificial intelligence (AI) model that has recently garnered significant attention due to its efficiency and cost-effectiveness. The China-based AI company built the open-sourced model using a mixture-of-experts (MoE) training on other leading models and an optimized architecture that only uses a subset of the model's parameters for each input. This reduces computational costs while performing on par with other leading large language models (LLMs) for simple use cases.

One of the key debates on DeepSeek is that the company claimed it took only 6 mn USD[1] to train the model, a fraction of the cost compared to other foundational models. This raises concerns about the sustainability of the current levels of AI spending in the industry. We believe the recent volatility in the technology sector, was a bit of an overreaction and view DeepSeek's incremental breakthrough as a net positive for AI innovation ahead.

- **DeepSeek is built on open source, not made from scratch**: DeepSeek's model was built on open source, leveraging collective expertise across other foundational models in the ecosystem. Moreover, tremendous computational

1 DeepSeek-V3 Technical Report

Sebastian Thomas
Voya Investment Management

James Chen
Voya Investment Management

Stephen Jue
Voya Investment Management

resources and investments were needed to build this collective expertise. The research and development for the MoE foundational models is likely not included in that 6 mn USD training figure and likely represents only the final training iteration.

- **Other competitors may not be far behind**: DeepSeek is a distillation model and distillation is not a new concept. Every model company has a "mini", distilled version of the bigger model. DeepSeek's model is more cost effective on a per-token (unit of intelligence) pricing perspective. This was also true of OpenAI o3 mini, which cost less than o1 mini. This was a first time a new model was priced below a prior equivalent model. Llama has a smaller distilled model that can be run locally as well. More efficient training has been a key focus across the AI ecosystem in order to bend the cost curve.

- **DeepSeek may have access to advanced NVIDIA GPUs**: Scale AI CEO Alexandr Wang suggested that DeepSeek may have been trained on a large cluster of advanced NVIDIA GPUs.[2] DeepSeek cannot talk about the use of advanced NVIDIA GPUs because of ongoing export restrictions. There may be reasons to be sceptical about the widely circulated narrative that DeepSeek was built on inferior NVIDIA chips.

- **Spending intentions from hyperscalers likely to remain elevated**: Note that DeepSeek

unveiled their open source LLM in December. Hence, hyperscalers have been aware of DeepSeek, as they have made their recent capital expenditure announcements that included Microsoft's 80 bn USD[3] and Meta's 60-65 mn USD[4] for 2025. We expect that most companies will reiterate their capital spending plans for AI when they report earnings over the next few weeks, which should help to calm down some of the concerns.

- **New use cases and broader adoption**: DeepSeek is an open-source model, so their breakthroughs have been shared with the broader AI developer community. The latest AI models still have limitations in complex reasoning and true understanding, with significant AI infrastructure required to support the trajectory of development. Next generation AI semiconductor chips are on the horizon and AI advancements are rapidly evolving. If models do indeed become much cheaper to train and run, this in turn should accelerate AI adoption and inference demand.

- **Consider the Jevons Paradox in AI**: The Jevons paradox is an economic principle where increased efficiency leads to even greater consumption of resources. The lower cost makes it more attractive to consumers and industries. We saw this play out when OpenAI introduced their ChatGPT pro subscription service

in 2024, as improved model efficiency caused consumption to vastly exceed expectations to the point where OpenAI was losing money on the subscription service. If DeepSeek's approach yields more efficient models, we may see even more demand for AI inferencing in the future.

Overall, while some believe that DeepSeek's progress could be an inflection point for the AI theme, we believe it is more of an incremental step forward in the AI innovation journey and not a game changing event. The industry is rapidly evolving, with major investments and innovations still needed to drive the progress towards artificial general intelligence. As AI models become smarter and more efficient, adoption should accelerate and drive more usage across a new class of AI applications. The stakes are high, as advancements in AI can lead to significant breakthroughs across industries from technology to defence, healthcare and finance. And we do not foresee any country or leading innovative company not investing for the opportunities ahead. We are still in the early days of the AI era and there is a lot more potential to be unlocked in the future.

2 Scale AI CEO Alexandr Wang on U.S.-China AI race (CNBC)
3 The Golden Opportunity for American AI – Microsoft On the Issues
4 Zuckerberg sets Meta's AI targets for the year, expects to spend $60 billion on growth

January 2025