

JAG Growth Equity Thematic Insights

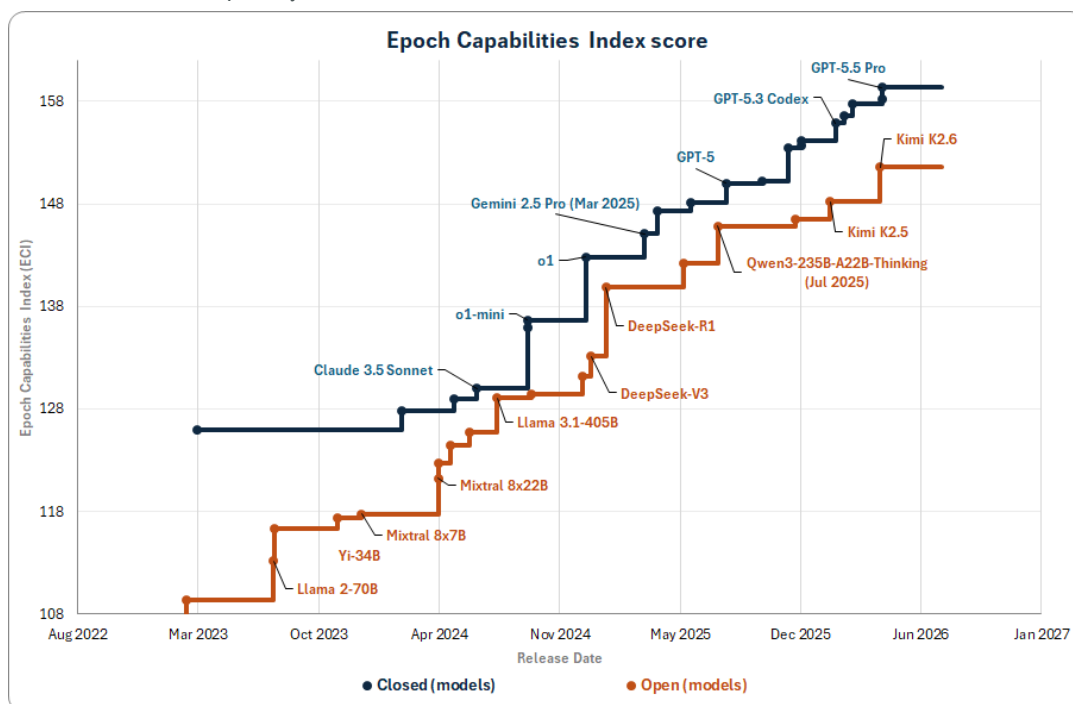
Total reading time = 5 minutes

Large Language Models and the Race to Cheaper Intelligence

A year after DeepSeek drew investor attention to lower-cost frontier AI models, Chinese AI startup Moonshot reignited the debate with the launch of its newest large language model (LLM), Kimi K3. While the model still trails the strongest proprietary offerings from Anthropic and OpenAI, the performance gap has narrowed significantly and Chinese LLMs continue to offer meaningful cost advantages. This raises an important question for investors: what does lower-cost AI mean for the broader AI ecosystem and where will the most value accrue?

Before diving into the broader investment implications, we think it is worth clarifying the terminology around AI models. Models such as DeepSeek and Kimi are often referred to as “open source,” but they are more accurately described as “open weight.” The distinction is important. Unlike open-source software which allows developers to access and modify the source code, open-weight AI models only provide access to the trained parameters (i.e., “weights”), not to the model’s underlying training data or methodology. This makes recreating open-weight models from scratch very difficult, if not impossible. Open-source AI models also exist, but their performance currently lags both open-weight and “closed” models like Claude or ChatGPT, which do not publicly release their weights or training data.

Figure 1: Open vs Closed Model Capability Index



Source: Epoch AI, JAG Research

Beyond lower costs, open-weight models also offer greater flexibility, giving users more control over where and how to run them. Companies can self-host models using their own equipment, keeping sensitive data within a controlled environment, and reducing third-party exposure. On the other hand, these models also come with several shortcomings. Real-world model performance can be inconsistent, and enterprise support is generally weaker than with closed-model providers. Self-hosting increases operational complexity because it requires a full-supporting AI infrastructure stack. However, much of this burden can be offloaded to one of the public clouds (Amazon Web Services, Azure, Google Cloud). Finally, while self-hosting meaningfully reduces data privacy concerns, it still does not eliminate geopolitical and regulatory risks. Embedded political biases remain a concern with Chinese open-weight models, and the US government could potentially restrict the use of future Chinese models on national security grounds.

What does the rise of open-weight models mean for investors? Implications differ across the AI value chain, which could create beneficiaries while simultaneously increasing competitive pressures in other areas. **We believe compute and AI infrastructure providers should remain the prime beneficiaries regardless of how the open vs closed model debate evolves.** All models require significant computing infrastructure, and cheaper models will further expand the scope of economically viable AI use-cases. The net effect is greater demand for inference – and with it, more demand for semiconductors, networking equipment, and memory/storage. The Kimi K3 release itself illustrated the point: the company was forced to pause new subscriptions just days after launch due to supply constraints. If inference demand continues to grow at or above the current pace, it is unlikely that computing supply will catch up to demand anytime soon.

Application-layer software companies could also be beneficiaries of open-weight models. Given the fact that AI models represent additional input costs (and potentially tighter margins) for software companies, competition between open and closed models could drive costs down. We have already seen companies actively incorporate open models into their workflow. For example, Duolingo (DUOL), a language learning platform, was able to reduce the cost of its video call feature from \$0.30 to \$0.01 per call after switching to open-weight models. Spotify (SPOT), a leading audio streaming platform, allows its engineers to “switch models mid-task and route every job to the best available price performance, including open-source models” to better control costs. As open-weight models continue to improve, we expect this trend to accelerate.

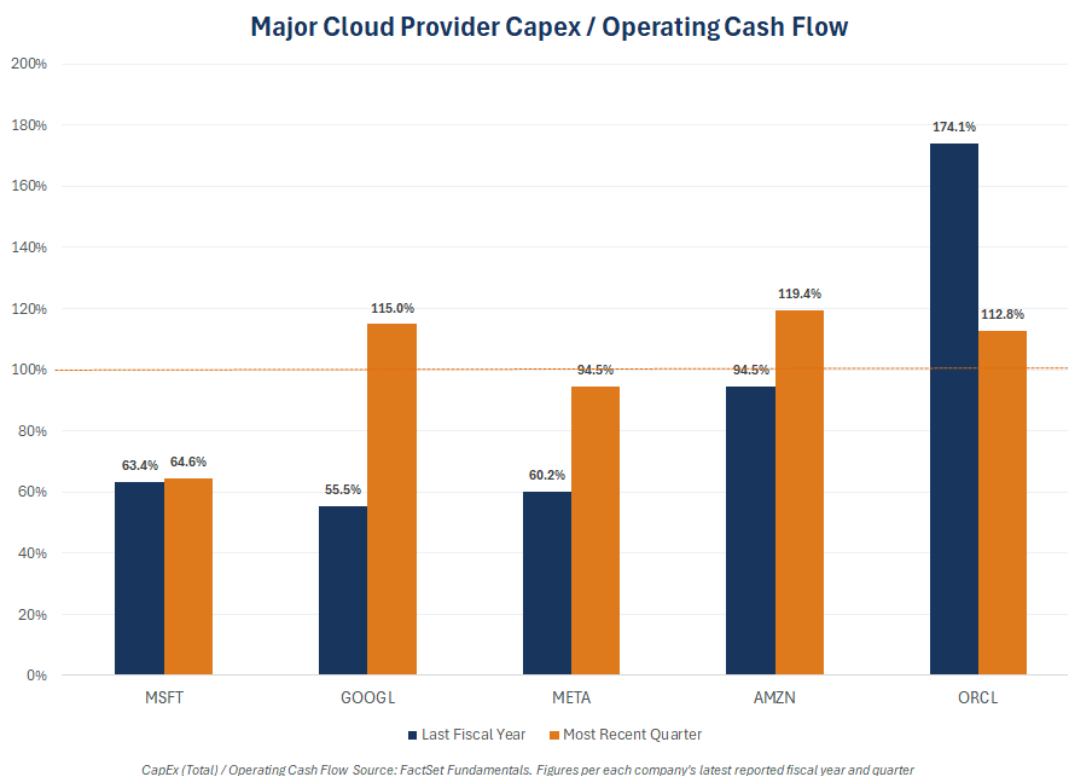
More open models should be a net positive for major cloud providers such as Microsoft, Amazon, and Google. While these companies all have their internal models that will face growing competition, their cloud businesses should significantly benefit from broader adoption of open models. Given the complexity and high cost of building and maintaining the AI infrastructure, most enterprises will likely outsource their computing needs to public cloud infrastructure rather than build and operate their own. As a result, a greater adoption of open models could further accelerate public cloud growth.

For model providers such as OpenAI and Anthropic, open models create a more direct competitive threat. Spotify is a good example – models are becoming more interchangeable, and workloads that previously would have exclusively gone to closed-model providers can now be routed to open alternatives. Additionally, while OpenAI and Anthropic continue to lead on performance, there are a growing number of tasks where the gap is small enough that some customers are comfortable using a slightly weaker model in exchange for cost saving. Longer term, success for closed-model providers will likely depend on their ability to maintain or expand their performance lead and drive differentiation throughout the entire ecosystem. This includes agents, connectors, workflow orchestration, and other proprietary tools that increase switching costs and make the underlying model just one part of the equation.

Despite our constructive view, open model improvements also carry risks for investors. If OpenAI and Anthropic are forced to materially cut their pricing, lower profitability/margins could eventually slow their capex growth. This could create headwinds for AI hardware providers. Another risk is that Jevons Paradox, where efficiency gains drive higher total usage, does not hold and computing needs fall faster than AI adoption grows – ultimately weakening demand for computing. More broadly, major cloud providers are spending

hundreds of billions on AI infrastructure buildout. This is significantly pressuring their cash flow generation (Figure 2), with some of them turning to debt and equity markets for additional financing. If monetization takes longer than anticipated or growth rates slow, this could force these companies to materially reduce spending.

Figure 2: AI Capex and Cash Flow Generation



Source: FactSet, JAG Research

That said, we currently see little evidence that these risks are materializing. Cloud growth accelerated across the board with Amazon Web Services, the world's largest cloud provider, reporting +37% y/y growth, up from +28% in the prior quarter. Microsoft's Azure business grew +43%, while Google Cloud continued to grow the fastest (+82%), albeit from the smallest base among the three. Despite this rapid growth, management teams consistently emphasized that AI demand still exceeds supply, and they will continue to aggressively invest in capacity expansion tied to record backlogs and strong visibility into demand growth. In terms of model providers' pricing dynamics, OpenAI recently cut prices for certain models amid intensifying competition and efficiency/cost improvements. Overall, current trends suggest that Jevons Paradox likely holds for now, while intensifying competition has yet to materially disrupt frontier-model economics, which is supportive of continued heavy investment in AI infrastructure.

– JAG's Growth Equity Research Team

Disclosures

These comments were prepared by the staff of JAG Capital Management, LLC, an SEC-registered investment adviser. The information herein was obtained from various sources including but not limited to FactSet, Bloomberg, Reuters, Standard & Poor's, Epoch AI, ChatGPT, Claude, and the United States Bureau of Labor Statistics, and believed to be reliable; however, we do not guarantee its accuracy or completeness. The information in this report is given as of the date indicated. We assume no obligation to update this information, or to advise on further developments relating to securities discussed in this report. The opinions expressed are those of the adviser listed above as of the date of this report and are subject to change without notice. The opinions of individual representatives may not be those of the Firm. Additional information is available upon request.

The information contained in this document is prepared and circulated for general information only. It does not address specific investment objectives, or the financial situation and the particular needs of any recipient. Investors should not attempt to make investment decisions solely based on the information

contained in this communication as it does not offer enough information to make such decisions and may not be suitable for your personal financial circumstances. You should consult with your financial professional prior to making such decisions. For institutional investors: JAG Capital Management, LLC, has a reasonable basis to believe that you are capable of evaluating investment risks independently, both in general and with regard to particular transactions or strategies. For institutions who disagree with this statement, please contact us immediately.

Past performance should not be considered indicative of future performance. Any investment contains risk including the risk of total loss.

This document does not constitute an offer, or an invitation to make an offer, to buy or sell any securities discussed herein. J.A. Glynn & Co., JAG Capital Management, LLC, and its affiliates, directors, officers, employees, employee benefit programs and discretionary client accounts may have a position in any securities listed herein.

About JAG

JAG Capital Management (JAG) actively invests for institutions and individuals in highly selective, customizable, and nimble equity and fixed income strategies. JAG is a boutique, independent, employee-owned investment management firm in St. Louis.



1610 Des Peres Road, Suite 120
St. Louis, MO 63131

