Can the DeepSeek wind fill the sails of cloud software companies?

Published 12 February 2025

Baoqi Zhu

Associate Director, Quantitative Research & Multi Asset Solutions

Elvira Kuramshina

Associate Director, Quantitative Research

Key Takeaways

- DeepSeek's Mixture of Experts and pure reinforcement learning (RL) improve efficiency, lowering compute costs without sacrificing scale.
- The "catfish effect" forces artificial intelligence (AI) providers to cut prices, making AI more affordable for Software-as-a-Service (SaaS) firms.
- Major cloud platform integration with DeepSeek R1 eases privacy and regulatory concerns, accelerating AI adoption across software companies.
- Cloud software companies are poised to benefit from AI advancements through product differentiation, cost efficiencies, and new revenue streams.
- A shift toward agentic AI and vertical AI is reshaping cloud software, enabling autonomous systems and domain-specific AI applications that offer a competitive edge.

DeepSeek has been making headlines since the eve of the Year of the Snake (24/01/2025), with financial markets reacting sharply and mixed. Meanwhile, discussions around data privacy, intellectual property, and geopolitical tensions have followed. However, we would like to take a step back from these broader debates and focus on the technical advancements of DeepSeek's models and their potential impact on the software sector.

What innovations did DeepSeek develop?

DeepSeek's artificial intelligence (AI) model is built on the Transformer architecture, but it incorporates multiple optimisations in design and engineering to improve efficiency. Rather than introducing entirely novel methods, DeepSeek combines existing innovations and addresses practical challenges—such as limited computing power—to balance theoretical advances with strong engineering execution.

In particular, we believe that DeepSeek's models, DeepSeek V3 and DeepSeek R1 Zero, feature two key breakthroughs:

Mixture of Experts (MOE) with auxiliary-loss-free strategy: DeepSeek-V3 divides the model into multiple 'expert' modules to distribute computational workload efficiently. During training, different experts run on different GPUs, improving efficiency. During inference, only 37B active parameters are used per token instead of the full 671B, significantly reducing compute costs. A common challenge with MoE is workload imbalance among expert modules; DeepSeek addresses this using relevance-based routing instead of auxiliary loss functions, ensuring all tokens are maintained during training and inference.

Reinforcement Learning without human feedback: DeepSeek-R1 Zero is a reasoning model like GPT-o1, and it demonstrates that pure reinforcement learning (RL) can develop reasoning capabilities without relying on Reinforcement Learning from Human Feedback (RLHF). Most large language models (LLMs) involve human labelling and feedback through supervised fine-tuning (SFT)[1] and RLHF. However, DeepSeek R1 Zero replaces SFT and RLHF with pure RL without human feedback.

Traditional RL has a high computational cost and limited scalability to diverse tasks. By introducing an innovative method called Group Relative Policy Optimisation (GRPO), DeepSeek R1 Zero addresses these challenges, plus, DeepSeek's progress indicates that a pure RL method could be an alternative exploration of how to continuously improve AI models given that the high quality, human generated data is finite and might be used up. This enhances the sustainability of model capability improvements.

The impact on software companies

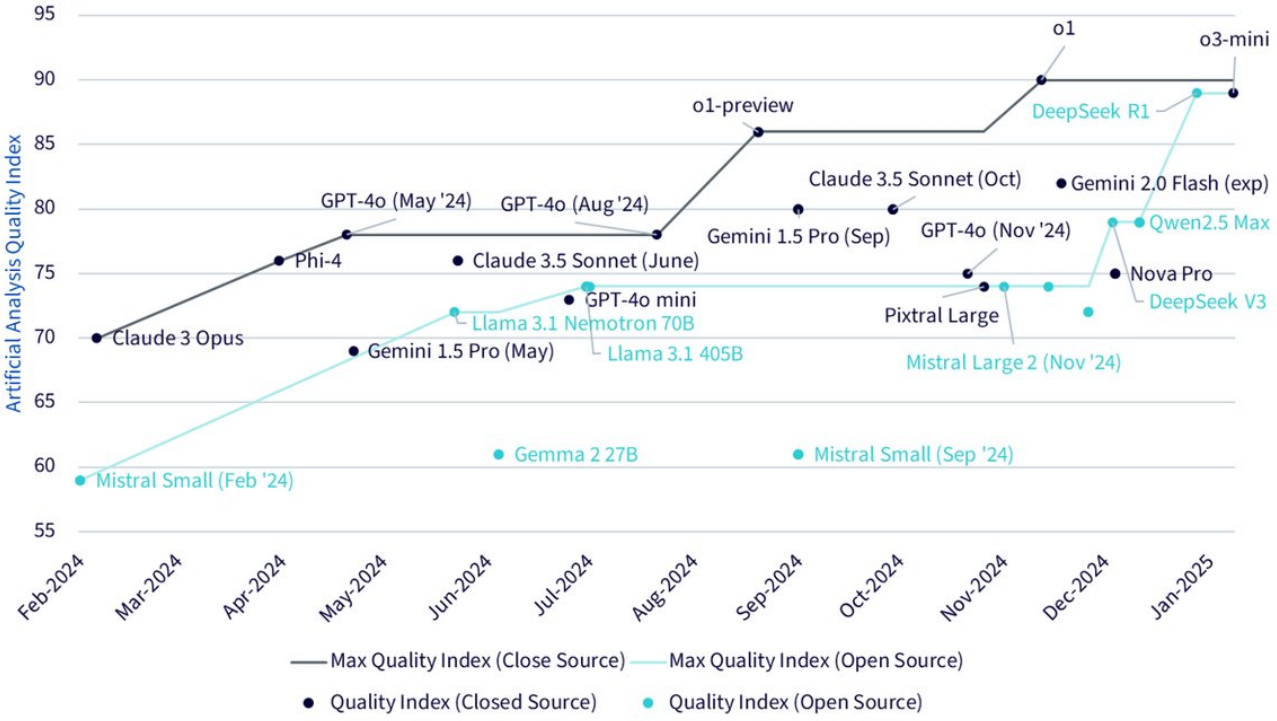
Since the release of ChatGPT, software investors have wrestled with a big question: how much value does AI add to SaaS companies? The developments of AI raised concerns over the sustainability of software growth—particularly whether AI could eventually replace traditional software. In addition, integrating AI models into SaaS offerings has introduced new usage costs, sparking doubts about potential margin pressures. However, recent developments suggest a more favourable outlook, easing some of these concerns.

A key driver of this shift is DeepSeek, whose open-source models rival the performance of leading large language models but are cheaper to use. This affordability lowers the barrier for SaaS firms to add AI features, letting them integrate AI into businesses with relatively lower expenses. In addition, leading cloud platforms including Microsoft Azure, AWS and Google Cloud have already added DeepSeek R1 to their services and this addresses privacy and regulatory concerns by deploying outside of China. Several US companies have already embraced DeepSeek. Perplexity, known for its AI-driven Q&A services, integrated the DeepSeek R1 model in their search engine. Cerebras, Groq also joined the trend. Meanwhile, ZoomInfo has expressed an interest in purchasing access to DeepSeek R1 through US cloud platforms[2].

DeepSeek's competitive pricing may trigger a 'catfish effect', prompting other model providers to release

more cost-effective and cheaper models. For example, OpenAI has already launched GPT-o3 mini at a very low-price level, with an output cost per 1 million tokens of only $4.4, around 63% lower than GPT-o1 mini's pricing. This competition benefits software companies, allowing them to adopt AI without eroding profit margins. In addition, DeepSeek has published its optimisation methods, which boosts open-source communities and may further narrow the gap between open-source and closed-source models. Over time, such openness should sustain competition and keep AI costs manageable.

Figure 1: The performance gap between closed-source models and open-source models is narrowing



Source: Artificial Analysis. Artificial Analysis Quality Index refers to average result across our evaluations covering different dimensions of model intelligence. Currently includes MMLU, GPQA, Math & HumanEval. OpenAI o1 model figures are preliminary and are based on figures stated by OpenAI. Data as of 31 January 2025. **Historical performance is not an indication of future performance, and any investments may go down in value.**

Figure 2: DeepSeek's APIs are relatively affordable compared to major AI models

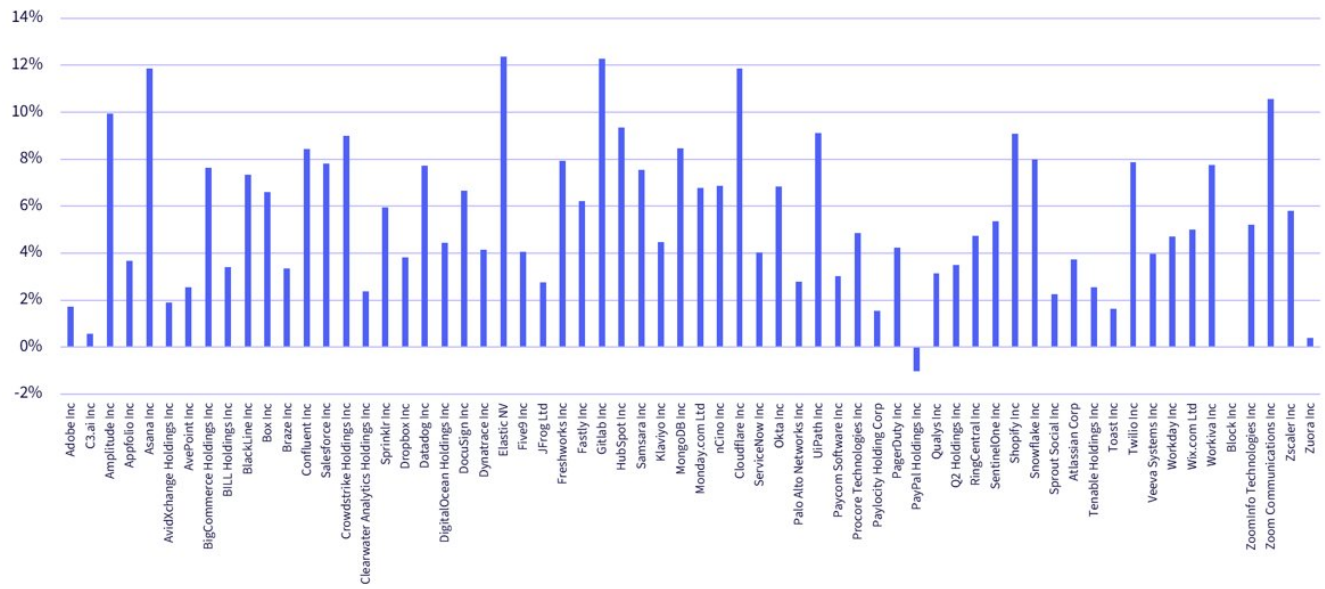
Company	Model	Input Cost Per 1M Tokens ($)	Output Cost Per 1M Tokens ($)	Artificial Analysis Quality Index
DeepSeek	DeepSeek R1	0.55	2.19	89.00
DeepSeek	DeepSeek V3	0.27	1.10	79.00
OpenAI	GPT-o1	15.00	60.00	90.00
OpenAI	GPT-o3 mini	1.10	4.40	89.00
OpenAI	GPT-o1 mini	3.00	12.00	84.00
OpenAI	GPT-4o	2.50	10.00	75.00
Anthropic	Claude 3.5 Sonnet	3.00	15.00	80.00
Google	Gemini 2.0	NA	NA	82.00
Google	Gemini 1.5 Pro	1.25	5.00	80.00

Source: Bloomberg, DeepSeek, OpenAI, Anthropic, Google, Artificial Analysis. The API prices are provided by model owners and may differ from prices offered by third party providers. Artificial Analysis Quality Index refers to average results across our evaluations covering different dimensions of model intelligence. Currently includes MMLU, GPQA, Math & HumanEval. OpenAI o1 model figures are preliminary and are based on figures stated by OpenAI. Data as of 31 January 2025. **Historical performance is not an indication of future performance, and any investments may go down in value.**

Can cloud software ride the DeepSeek wave?

The BVP Nasdaq Emerging Cloud Index selects pure-play opportunities within the cloud computing landscape focusing on those outpacing the broader market in revenue growth. Using this index as a proxy for investor sentiment, we assessed market reactions to DeepSeek's release on 27-28 January. Excluding PayPal, all companies reacted positively, with a median return of 4.9% over the period (see Figure 3). Since no earnings reports were released during this timeframe, we attribute this reaction mostly to DeepSeek's announcement. Further, the Federal Open Market Committee (FOMC) meeting on 28-29 January tempered sentiment toward the end of the week, bringing the median five-day return down to 2.7%.

Figure 3: Returns of cloud computing companies following DeepSeek's release



Over the past two years, cloud software companies have not been viewed as direct AI beneficiaries, unlike AI infrastructure players and model developers. Their role in the AI revolution was seen as downstream, requiring them to first integrate AI into their products before capturing value. Additionally, concerns arose over AI-related costs impacting margins, especially in the context of lower pricing power amidst macroeconomic headwinds and constrained software budgets.

An era of more affordable AI models, heralded by DeepSeek's release, marks a potential turning point, offering cloud software companies opportunities for product differentiation, cost efficiencies, and new revenue streams. Recent trends suggest a strategic shift in the cloud software space, with companies increasingly focusing on agentic AI (autonomous systems that perform tasks with minimal human intervention) and vertical AI (highly specialised AI models in a specific industry or domain, for example, healthcare, legal, finance). The emergence of AI-native cloud companies further underscores this evolution, positioning cloud providers to benefit more directly from AI advancements.

DeepSeek's release sparking a new wave of innovations in the AI space is set to accelerate AI adoption. As AI adoption accelerates, companies are poised to increase their spend on software with rising demand to leverage AI technology across the board. In the shorter term, a more hawkish monetary policy in the US presents a headwind for cloud software providers, but earnings releases throughout February and early March might offer a line of sight into companies' plans to capitalise on this AI-driven transformation of the space.

[1] In the following version, DeepSeek-R1, SFT was applied in model training, but its focus was on improving the readability and coherence of the model's responses. The model's reasoning and decision-making capabilities were primarily refined through reinforcement learning (RL).
[2] The Information: https://www.theinformation.com/articles/deepseek-attracts-surge-of-business-users.

Important Risks Related to this Article

Important Information

Marketing communications issued in the European Economic Area ("EEA"): This document has been issued and approved by WisdomTree Ireland Limited, which is authorised and regulated by the Central Bank of Ireland.

Marketing communications issued in jurisdictions outside of the EEA: This document has been issued and approved by WisdomTree UK Limited, which is authorised and regulated by the United Kingdom Financial Conduct Authority.

WisdomTree Ireland Limited and WisdomTree UK Limited are each referred to as "WisdomTree" (as applicable). Our Conflicts of Interest Policy and Inventory are available on request.

For professional clients only. The information contained in this document is for your general information only and is neither an offer for sale nor a solicitation of an offer to buy securities or shares. This document should not be used as the basis for any investment decision. Investments may go up or down in value and you may lose some or all of the amount invested. Past performance is not necessarily a guide to future performance. Any decision to invest should be based on the information contained in the appropriate prospectus and affer seeking independent investment, tax and legal advice.

The application of regulations and tax laws can offen lead to a number of different interpretations. Any views or opinions expressed in this communication represent the views of WisdomTree and should not be construed as regulatory, tax or legal advice. WisdomTree makes no warranty or representation as to the accuracy of any of the views or opinions expressed in this communication. Any decision to invest should be based on the information contained in the appropriate prospectus and affer seeking independent investment, tax and legal advice.

This document is not, and under no circumstances is to be construed as, an advertisement or any other step in furtherance of a public offering of shares or securities in the United States or any province or territory thereof. Neither this document nor any copy hereof should be taken, transmitted or distributed (directly or indirectly) into the United States.

Although WisdomTree endeavours to ensure the accuracy of the content in this document, WisdomTree does not warrant or guarantee its accuracy or correctness. Where WisdomTree has expressed its own opinions related to product or market activity, these views may change. Neither WisdomTree, nor any affiliate, nor any of their respective officers, directors, partners, or employees accepts any liability whatsoever for any direct or consequential loss arising from any use of this document or its contents.