

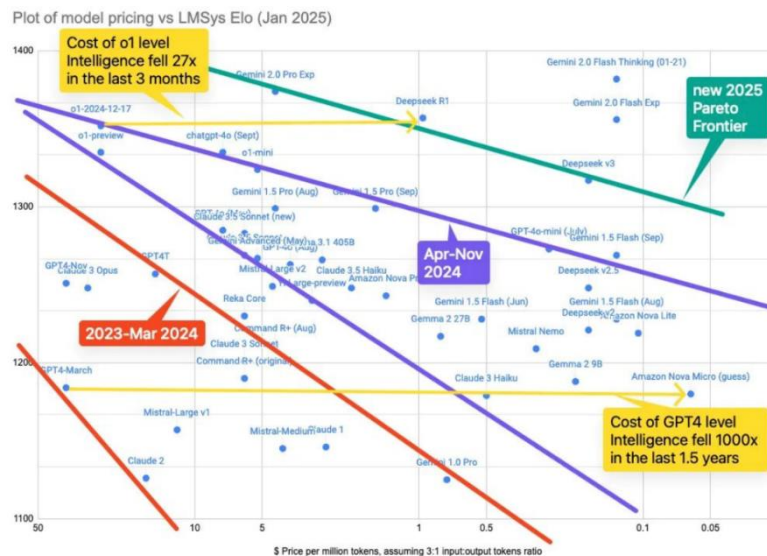
## Alpha Helix Asset Management 投資備忘錄 (七十一)

2025.01.31

### DeepSeek 是否會降低 CSP 對 AI 硬體的需求?

#### DeepSeek 公布開源模型 R1，NVDA 等半導體股價重挫

中國量化基金幻方的子公司，AI 新創 DeepSeek，在去年 12 月和今年 1 月分別發布了 DeepSeek V3 和 R1 模型，並宣稱 V3 和 R1 模型分別對標 OpenAI 的前沿模型 4o 和 o1，但訓練成本遠低於 OpenAI，更令人驚訝的是，DeepSeek R1 價格僅是 OpenAI o1 的 3%-4%。



(圖一)每個新世代的模型集群 (由紅向上至綠線)，成本下降的程度都更為顯著，顯示 AI 技術前緣的不斷突破。

source: @swyx on X

該新聞在 1 月 25, 26 日在社群平台 X 引起廣泛討論，並激起矽谷工程師和投資者的擔憂，認為美國 AI 技術已被中國超越，並過度投資 AI 的硬體設備。恐慌情緒在 1/27 周一開盤時爆發，當日 NVDA/TSM/ AVGO 皆下錯 13%-17%不等，其餘 AI 基礎建設股更下挫餘 20%。

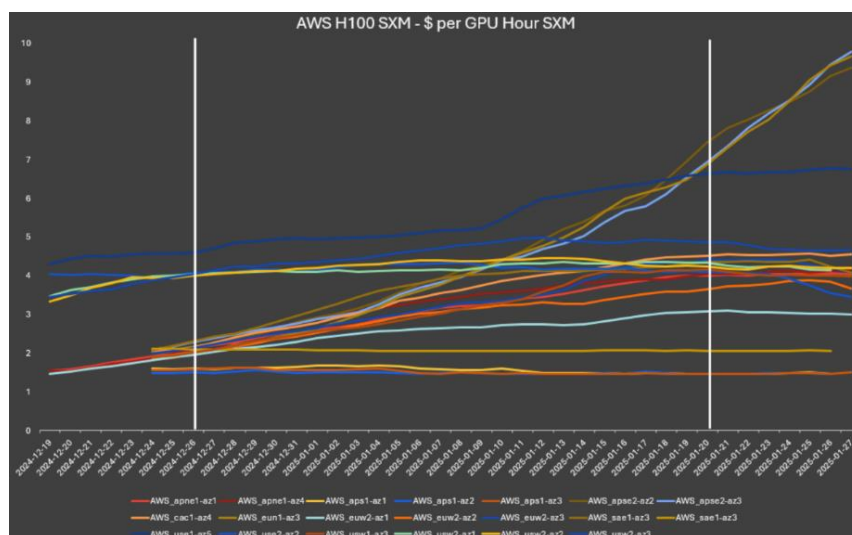
#### DeepSeek 的技術對整個產業代表了什麼? CSP 真的過度投資 AI 硬體設備了嗎?

我們認為，市場的下跌實為過度恐慌。DeepSeek 確實推進了技術前緣，提升了 AI 計算系統的使用效率，有效降低了當前模型的運行成本，但這並不會導致 CSP 減少對 AI 的硬體投資。

讓我們先回顧關於半導體將下行的論述，持悲觀看法的投資人認為，因為 DeepSeek 有效提升了技術前緣，原先要花費在同一個大小上的算力需求大幅下降，過去的 CSP 的投資被證明有更有效的替代方案，因此應被視過度花費。

然而，我們的論述與悲觀論述正好相反，我們的論點有二：

1. **Jevon' s Paradox**：DeepSeek 讓技術前緣大幅上升後，每單位算力的使用效率提升，降低了 LLM 的使用成本，原先因為使用 o1 模型成本過高而卻步的使用者，現在能用更便宜的成本就享受到優秀的模型能力，這會導致模型的使用量增加。在經濟學上這個概念被稱為 Jevon' s Paradox，當一個產品的需求彈性很高時，該產品的價格降低將刺激更多的需求，使得整體市場的份額(降低的價格和激增的需求的乘積)增加。
2. **更快速的科技迭代，致使競爭者（即 CSP）更迫切地提高資本支出以維持領先地位**：細看 DeepSeek 所釋出的論文內容，該公司雖然研發出了新技術，但這些技術並沒有顛覆既有的 AI Scaling Law (想複習 AI Scaling Law 的讀者可以閱讀和奕在去年發布的文章)。反而，DeepSeek 展現的新技術更證明了 AI Scaling Law 中，Post-Training Scaling Law 和 Test-time Scaling Law 的延續性及潛力。這會加速 CSP 原有的軍備競賽，投入更多的資本支出，因為其他的參賽者現在知道，投入更多的算力，能夠得到更有效率，表現更好的 LLM。



(圖二) DeepSeek 推出 V3 (左線) 和 R1 (右線) 後，在 AWS 上 H100 的租賃價格不減反增，且幅度極大，顯示對於 AI 硬體需求旺盛。 source: Semianalysis

綜上，我們認為，因為 DeepSeek 的橫空出世，CSP 及其他企業對於算力的需求反而會更加飢渴。由上圖可以觀察到，Nvidia H100 的租賃價格已初現端倪，自從 DeepSeek 推出 V3 和 R1 後，H100 的租賃價格不減反增，表示需求的增加高過算力的供給速度，我們認為這樣的情況將隨著 AI 技術前緣的持續推進而持續發展，這對於 AI 硬體供應商是很好的消息。