

〔投資研究報告〕

2025/11/13

AI 狂潮：三兆級長跑， 誰賺得回來、誰會先撐不住？

目錄

- 一、收入端：AI 需求真有那麼大？還是投資已經走在前面？
- 二、成本端：推理單價下降，卻擋不住電力與基建的成本地板
- 三、資金結構：誰用的是便宜錢、誰撐在槓桿前線？
- 四、用三個問題審視任何一檔「AI 概念股」

從 GPU 搶購戰，走向回收期大考驗

2025/11/13 新光投顧

短短兩三年，生成式 AI 從新玩具變成資本市場的主角。

NVIDIA 的市值衝上天已是耳熟能詳，更關鍵的是背後一整條「算力產業鏈」——超級雲端業者 (hyperscalers)、AI 晶片、伺服器、資料中心、電力及冷卻基建——正以近乎前所未見的速度堆疊資本支出。

摩根士丹利估計，Amazon、Google、Microsoft、Meta 等幾家超級雲的資本支出，在 2025 年合計將逼近 3,000 億美元，而推動成長的最主要動能，正是 AI 相關投資。同一時間，OpenAI 等模型公司不只大舉購買或租用 GPU，也正在簽下跨雲端、多年的大型合約：

- OpenAI 已與 Amazon Web Services (AWS) 簽署 七年、380 億美元的雲端服務協議，確保 2026 年前可取得「數十萬張」NVIDIA GPU 與相關運算資源。
- 與 Oracle 方面，OpenAI 透過 Stargate 合作，計畫在美國新增 4.5GW 資料中心容量，使 Stargate 建設中的總量超過 5GW、支撐逾 200 萬顆 AI 晶片運作；部分媒體與產業分析甚至推估，相關雲端與基建合約的金額可能達每年數百億美元級距，但雙方對具體金額並未正式證實。

換句話說，這已經是一場「兆美元級別」的長期押注。

然而，對一般投資人來說，關鍵問題其實只有一句：

這些錢，未來真的賺得回來嗎？

如果賭錯，哪一個環節/哪些股票會先跳樓？

要回答這個問題，我們只需要冷靜地從三個面向深入：

(1) 收入能見度與需求品質 (2) 成本與電力壓力 (3) 資金結構與槓桿敏感度。

目錄

- 一、 收入端：AI 需求真有那麼大？還是投資已經走在前面？
- 二、 成本端：推理單價下降，卻擋不住電力與基建的成本地板
- 三、 資金結構：誰用的是便宜錢、誰撐在槓桿前線？
- 四、 用三個問題審視任何一檔「AI 概念股」

一、 收入端：AI 需求真有那麼大？還是投資已經走在前面？

1. 超級雲：AI 支撐成長，但市場開始問「賺多少」而不是「花多少」

從財報來看，AI 確實正在支撐雲端業務：

- Microsoft、Google、Amazon 在最新季度皆指出，AI 工作負載是雲端成長的重要來源之一。
- 零售、金融、製造等產業客戶的 AI 試點專案不斷增加，從客服、行銷內容到程式開發自動化，全面開花。

問題是——收入成長速度，未必能跟上資本支出膨脹的速度。

摩根士丹利與其他券商的報告近來開始轉向提醒：雲端整體成長仍健康，但傳統 IaaS / PaaS 的成長有放緩跡象；若 AI 相關收入無法在毛利率與規模上快速補位，市場對「幾千億美元 AI CapEx」的耐心，終究會有極限。

換句話說，現在的市場關注焦點已經從「雲端收入有沒有增加」轉移到「AI 的每一塊錢投資，到底能多賺錢」。

2. OpenAI：收入成長驚人，但算力承諾更驚人

OpenAI 是這一輪 AI 投資的焦點。從最新公開數字來看：

- 路透與 The Information 等報導指出，截至 2025 年中，OpenAI 年化營收已達約 100–120 億美元，較 2024 年底幾乎翻倍，且每月營收約在 10 億美元級。

這個規模對一家不到十年歷史的公司來說已是驚人，但若放在它的「算力承諾」與「資本計畫」背景下看，故事又完全不同：

- 與 AWS 的 380 億美元七年合約，保證了龐大的 GPU 與伺服器供給。
- 與 Oracle 的 Stargate 計畫，預期建成後將提供超過 5GW 的 AI 專用資料中心容量，配備約 200 萬顆晶片，用電規模相當於數百萬戶家庭。
- 再加上與 Microsoft 長期合作、對 GPU 供應商（如 NVIDIA、AMD）的合約與授權支出，外界普遍認為 OpenAI 未來數年在算力上的承諾，「遠高於」目前每年十幾億美元的營收規模。

這不代表一定會出事，反而是把商業邏輯寫得非常清楚：

先用長約鎖住未來十年的算力供給，再用產品、訂閱與企業方案，努力把這些算力「變現」。

換成投資人的語言：

支出與承諾已經大幅走在收入前面，這本質上是一場「賭未來現金流」的大型槓桿實驗。

3. 企業導入：從鋪天蓋地的試驗，走向 ROI 的殘酷分化

要讓上游的算力投資真正賺錢，企業端導入是關鍵。但目前看到的是一個「結果高度分化」的世界：

- 某些應用非常亮眼：例如客服自動化、程式碼輔助、內部知識問答，確實能節省工時或提高效能。
- 但也有不少專案在試驗一兩年後發現：模型成本偏高、導入後流程更複雜、或員工採納度不如預期。

這種分化意味著：

- ✧ **AI 不會像 4G 那樣「全面普及、一致受益」**
不同產業、不同公司之間的導入速度差距，可能非常大。
- ✧ **真正願意「簽長約、綁用量」的客戶，仍在形成中**
短期試用與小規模專案會很多，但這些對上游 GPU 與資料中心的「回收期」，幫助有限。

對算力供應鏈來說，需求確實存在，但是「轉換成穩定現金流的速度」尚未被證明足夠快。

二、成本端：推理單價下降，卻擋不住電力與基建的成本地板

表面上看，AI 成本好像在快速下降——多家模型與雲端廠商都在調降每 1,000 tokens 的價格，推出更便宜的中小型模型。但若把所有成本加總，故事其實更複雜。

1. 每 1,000 tokens 變便宜，不等於每一次使用變便宜

單看價目表，生成式 AI 服務的確在「降價」：

- 多家雲端與模型供應商推出更便宜的基礎模型，價格只及旗艦模型的一小部分。
- 大量企業客戶可以透過批量合約獲得更佳折扣。

但實務上，有幾個重要「逆風」正在同時發生：

- 上下文變長**：模型支援十萬、甚至百萬 tokens 的 context，單次請求處理量增加。
- 多模態普及**：圖片、影片、語音輸入比重上升，單次推理消耗的算力與記憶體遠高於純文字。
- 延遲與體驗要求提高**：金融、遊戲、即時互動場景，要求低延遲，難以無限度用大批次方式壓低成本。

因此即使「每 1,000 tokens 價格下降」，「每一次實際服務」的成本未必同步下降，甚至可能小幅上升。這也是為何大型雲端在談 AI 服務時，仍對毛利率保持謹慎口氣。

2. 電力與資料中心：AI 算力的真正成本地板

真正把整體成本「墊高」的，是電力與資料中心基礎建設。

國際能源總署（IEA）在 2025 年發布的報告指出：

- 到 2030 年，全球資料中心用電量預計將超過 945TWh，約為目前的兩倍，略高於日本目前一整年的全國用電量。
- AI 將是推動這個成長的最大單一因素之一。
- 在許多已開發國家，資料中心可能貢獻電力需求成長的 20% 以上。

對營運者而言，這代表三層壓力：

- i. **電價難以下修**：在能源轉型與供需吃緊的背景下，中長期電價向下空間有限。
- ii. **電網與變電建設需時間**：IEA 警告，如果電網升級和新電源建設跟不上，資料中心專案恐面臨延宕或限量接入。
- iii. **冷卻技術 CAPEX 上升**：高功耗 GPU 伺服器需要液冷、浸沒式冷卻等方案，初期投資明顯高於傳統風冷。

換句話說，就算 NVIDIA 每一代晶片效能提升、每瓦算力變強，電力與基建這個「底座」仍在抬高整體成本地板。

三、資金結構：誰用的是便宜錢、誰撐在槓桿前線？

同樣是砸在 AI 上的錢，「誰出的」差很大，風險位置也完全不同。大致可以分成三個層次來看。

1. 超級雲端：用自家現金流與投資級債，壓力主要在估值

以 Amazon、Microsoft、Google、Meta 為代表的超級雲，資金來源主要是：

- 自家既有業務（廣告、電商、雲端、企業軟體）的自由現金流
 - 信評在投資等級之上的公司債
- 在這樣的結構下，它們的 AI 資本支出雖然巨大，但：
- **償債風險極低**，短期幾乎不可能出現流動性危機
 - 真正壓力來自於：
 - EPS 在折舊與攤提之下是否被壓縮太久
 - 市場是否願意持續給予「高成長、高本益比」的估值

對投資人來說，這一層的風險，在於「估值重評」。

2. 新興 AI 雲與 Neocloud：用長約 + 高槓桿撬起高速成長

第二層是專注 AI 算力租賃的雲端新勢力，例如 CoreWeave 等。

以 CoreWeave 為例：

- 2023 年取得 23 億美元的債務融資，由 Magnetar、Blackstone 等機構領投。
- 之後又持續透過延遲提款定期貸款（DDTL）、循環信貸額度等工具，把可動用債務規模擴大到數十億美元，用來預先購買 GPU、建置專用 AI 雲基礎設施。

這類公司的資金結構有幾個共通特徵：

- i. **高度依賴長約保障現金流**：只要與 OpenAI 等大客戶的合約順利履行，利息與折舊壓力可以被吸收。
- ii. **對利用率極度敏感**：一旦 GPU 閒置率高於預期，固定成本會很快吃光利潤。
- iii. **股權與次順位債是第一道緩衝**：一有風吹草動，股價與高收益債就會明顯反應。

也就是說，如果 AI 收入成長不如預期，最先感受到「時間不夠用」的，往往是這些站在槓桿前線的 Neocloud。

3. 應用層與軟體公司：不先死於債務，而是死於毛利率

第三層是站在「應用端」的 AI 公司：

從新創 SaaS，到傳統軟體廠引入 AI 功能，幾乎清一色是「租用雲端算力」，再包裝成訂閱或方案賣給客戶。

這一層的風險不是銀行或債務，而是：

- **推理成本占營收比重太高，毛利率上不去**
- 在利率仍具水準的世界，資本市場對「燒錢換成長」的耐心大幅下降

當推理成本長期吃掉毛利，即使營收看似高速成長，也難以轉化為實質獲利，估值一旦重評，下修幅度會非常驚人。

四、用三個問題審視任何一檔「AI 概念股」

整理上述脈絡，未來 12–36 個月，AI 產業不太像會爆出「一次性全面崩盤」那種劇情，而比較像是不同層次、不同位置，依照「時間是否站在自己這邊」陸續分出勝負。

對一般投資人來說，可以用以下三個問題來審視任何一檔「AI 概念股」：

問題一：這家公司賺的是哪一段價值鏈？

- 若它是 **超級雲、領先晶片廠、關鍵零組件供應商**：
→ 受惠於整體算力擴張，只要 AI 投資不急煞，長期趨勢偏正向，風險在估值高低。
- 若它是 **Neocloud 或高槓桿 AI 雲**：
→ 要特別小心利用率、長約條款與債務到期結構，一旦成長低於預期，股價波動可能極大。
- 若它站在 **應用層**：
→ 核心問題是：**毛利率是否能在推理成本之上，形成穩定利潤空間**。

問題二：它的賺錢邏輯，需要哪幾件事「同時成立」？

幾乎所有 AI 投資故事，最後都可以被拆成三句話：

- i. 客戶願意持續買單，AI 確實提升生產力或創造新收入。
 - ii. 單位算力成本會隨技術進步與規模，逐步往下走，而不是被電力與基建成本吃回去。
 - iii. 這家公司有足夠的財務彈性，撐到這兩件事發生。
- 只要有一項被現實打折，回收期就會拉長，估值就得重算。

問題三：如果 AI 投資放慢，這家公司會在第幾輪感到痛？

- **第一輪就感受到痛**：
高槓桿、新興 AI 雲、營收高度集中於少數 AI 客戶者。

- **第二輪才感到痛：**
應用層公司，在「成長故事」退潮時，面臨估值壓縮與融資困難。
 - **最後才真正受影響：**
超級雲與關鍵硬體供應商，通常有多元業務與充沛現金流，調整節奏的空間較大。
-

結語：喊出「AI 泡沫論」之前，我們先看「時間與現金流」的長期賽跑

綜合目前可見的產業與金融數據，很難把這一輪 AI 投資直接類比為 2000 年那種「沒有實質生產力支撐」的純泡沫，也不像 2008 年那種「靠銀行高倍槓桿堆出來」的系統性危機。

更貼切的描述是：

這是一場由科技巨頭和雲端業者主導的「三兆美元長跑」。

勝負不在於誰先砸最多錢，而在於：

收入成長、成本下降與電力與基建條件，

能不能在有限時間內跑到一條交會點之上。

對一般投資人而言，不用去預測哪一天會有「AI 崩盤」，而是：

- 看懂每家公司在算力鏈中的位置
- 判斷它有沒有能撐過多輪景氣與技術迭代的現金流與財務結構
- 不被單一的「題材故事」牽著走，而是持續追蹤收入、成本與資本開支三條線的變化

AI 很可能會在未來十年，成為全球生產力與資本市場的核心主軸之一。

但在這條路上，**不是每一個站在牌樓下喊「AI」的公司都能走到終點；**

對投資人來說，真正重要的，是選到那些**既站在對的位置，又撐得夠久、財務結構不會先出問題**的少數贏家。