

PRE-IPO INVESTMENT RESEARCH

Anthropic 與 OpenAI Token Economics、算力合約、ASIC 與估值風險

以公司官方披露、SEC / 交易對手申報、一級媒體與 Stock LLM Wiki 為依據；逐筆區分 fact、company_comment、estimate、rumor 與 inference。

研究截止日
2026-09-01

研究範圍
AI 基礎設施 × 私募市場 × pre-IPO

資料規則
一列一 claim；每列附原文證據

投資結論先行

核心 thesis：兩家公司都不是傳統高毛利 SaaS；它們更像「高研發強度的模型 IP 公司 + 以長約租用的 AI 雲端」。Anthropic 的企業/API mix 與多雲晶片策略較利於近期單位經濟，OpenAI 的規模、自研 ASIC 與供應多元化提供更大長期作業槓桿，但也帶來更重的合約承諾、融資與執行風險。

20.5×

Anthropic 2026-05 私募估值 / 同期官方 run-rate revenue (inference)

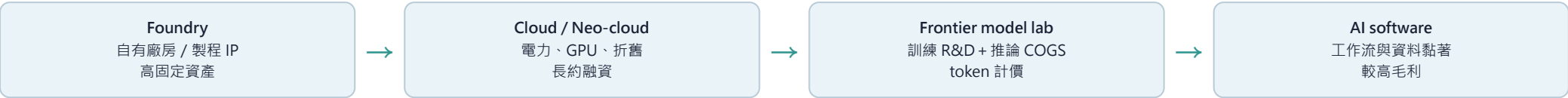
42.6%

OpenAI 2025 推算 gross margin；依外洩且經媒體核驗的 audited figures (inference)

\$250B

OpenAI 對 Azure 的 incremental services commitment (SEC fact)

商業模式位置



估值閱讀要點：run-rate revenue 不是 GAAP revenue；公司新聞稿中的「投資」也不等於已支出的 capex。大型合作中同時存在採購、投資、認股權與供應商融資，必須避免把所有金額直接相加。

可比結構 (截至截止日可得的最新披露)

類型 / 代表	公開毛利訊號	資本負擔	收入可見度	對 AI lab 的啟示
Foundry / TSMC	2Q26 gross margin 67.7%	2026 capex 指引 \$60-64B；資產自有	訂單與產能利用率驅動	若 AI lab 自建算力，資本密度會向製造業靠攏；但目前主要以外部長約承擔。
Cloud / Microsoft Cloud	FY26 gross margin 66%	資料中心與伺服器折舊；跨工作負載共用	企業合約與雲端消費	雲端可用資產池、企業分銷與較低資金成本，是 AI lab 暫時難複製的優勢。
Neo-cloud / CoreWeave	未以單一簡單比率呈現；COGS 含電力、租金、折舊	長約支撐資產層債務	2025 年 98% 以上收入來自 committed contracts	AI lab 的長約是 neo-cloud 融資抵押物，也把技術迭代與需求風險回傳給 lab。
AI software / ServiceNow	FY25 subscription gross margin 80%	主要為人員與第三方雲端；較輕	訂閱與續約	AI lab 若由 API 轉向高使用量訂閱，毛利未必接近 SaaS，因推論成本隨使用量上升。

完整 Datapoint 表

fact company_comment estimate rumor inference

ID	公司	主題	Datapoint / Claim	數值或說法	單位	日期	類型	來源	原文證據	信心	投資含義
A01	Anthropic	Tokenomics	Opus 4.8 標準 API 定價為 input \$5、output \$25。	\$5 / \$25	USD / 百萬 tokens	2026-05-27	fact	Anthropic List Prices；公司披露；URL	"BASE INPUT TOKENS \$5.00 ... OUTPUT TOKENS \$25.00"	高	輸出 token 單價為輸入 5 倍；output mix、推理長度是 revenue 與 COGS 的共同驅動。
A02	Anthropic	Tokenomics	Opus 4.8 cache hit 價格為 \$0.50，較標準 input 低 90%。	\$0.50；低 90%	USD / 百萬 tokens	2026-05-27	inference	Anthropic List Prices；公司披露 + 算術；URL	"CACHE HITS & REFRESHES \$0.50"	高	企業 agent 的重複 context 越高，單位服務成本越可壓低；實際利益取決於 cache hit rate。
A03	Anthropic	Tokenomics	Batch API 對 input 與 output 提供 50% 折扣。	50%	折扣	資料擷取 2026-09-01	fact	Anthropic Pricing Docs；公司文件；URL	"50% discount on both input and output tokens"	高	非即時工作負載可移至 batch；折扣也反映離峰排程與利用率管理的價值。
A04	Anthropic	營收	Series H 公告時，run-rate revenue 已跨過 \$47B。	>\$47B	USD / 年化 run-rate	2026-05	fact	Anthropic Series H；公司披露；URL	"our run-rate revenue crossed \$47 billion"	高	支持高估值的分母快速擴張，但不可等同全年已實現營收。
A05	Anthropic	毛利 / 獲利	2Q26 初步營收超過 \$11.5B，並呈正 adjusted operating income。	>\$11.5B；正數	USD / 季；調整後營業利益	2026Q2	estimate	Bloomberg News 文件報導；媒體見內部文件；URL	"reported positive adjusted operating income"	中	若 IPO 文件確認，將顯著降低「永續燒錢、折價」；仍需核對調整項與 stock-based compensation。
A06	Anthropic	毛利	截至截止日未找到可靠公開 GAAP gross margin。	未找到可靠公開資料	—	2026-09-01	fact	官方融資公告、公開報導與資料庫檢索；研究查無	"A confidential filing means there are no public numbers disclosed."	高	pre-IPO 定價不應把 API 高毛利敘事當成已審計毛利；S-1 是關鍵驗證點。
O01	OpenAI	Tokenomics	GPT-5.5 API 定價為 input \$5、cached input \$0.50、output \$30。	\$5 / \$0.50 / \$30	USD / 百萬 tokens	2026-04-23	fact	OpenAI model docs；公司披露；URL	"Input \$5.00 ... Cached input \$0.50 ... Output \$30.00"	高	輸出 token 價格為 input 6 倍；reasoning 長度與工具迴圈直接影響 monetization 與算力負荷。
O02	OpenAI	Tokenomics	GPT-5.5 cached input 較標準 input 低 90%。	90%	折扣	2026-04-23	inference	OpenAI model docs；公司披露 + 算術；URL	"Input \$5.00 ... Cached input \$0.50"	高	長 context agent 若能有效重用 prefix，推論毛利可改善；cache write 與 GPU-local storage 仍有成本。
O03	OpenAI	Tokenomics	Batch API 在 24 小時內完成，價格折扣 50%。	50%；24h	折扣；時限	資料擷取 2026-09-01	fact	OpenAI API Reference；公司文件；URL	"returns completions within 24 hours for a 50% discount"	高	離峰排程能提高設備利用率，但折價會把部分效率回饋給客戶。
O04	OpenAI	營收	外洩 audited statements 顯示 2025 revenue 為 \$13.07B。	\$13.07B	USD / 年	2025	estimate	Ars Technica / FT 核驗；媒體報導外洩審計文件；URL	"revenue growing from \$3.7 billion ... to \$13.07 billion"	中	比 run-rate 更適合作歷史倍數分母，但公司未正式發布，IPO 前仍需 S-1 確認。
O05	OpenAI	成本	同批文件顯示 2025 cost of revenue 為 \$7.5B。	\$7.5B	USD / 年	2025	estimate	Ars Technica / FT 核驗；媒體報導外洩審計文件；URL	"cost of revenue ... increased ... to \$7.5 billion"	中	顯示 serving cost 高於典型 SaaS；推論效率與訂閱用量限制是毛利關鍵。
O06	OpenAI	毛利	以 \$13.07B revenue 減 \$7.5B COGS 推算，2025 gross margin 約 42.6%。	約 42.6%	毛利率	2025	inference	同 O04/O05；媒體核驗數字 + 研究計算；URL	"revenue ... \$13.07 billion ... cost of revenue ... \$7.5 billion"	中	遠低於成熟 AI software 的 75-80% 毛利區間；但產品 mix 與效率改善可能快速改變。
O07	OpenAI	R&D	外洩文件顯示 2025 R&D 為 \$19.18B。	\$19.18B	USD / 年	2025	estimate	Quartz / FT 核驗；媒體報導外洩審計文件；URL	"\$19.18 billion in research and development spending"	中	訓練 compute 多落在 R&D 而非 COGS；只看 gross margin 會低估現金消耗。
O08	OpenAI	獲利	外洩文件顯示 2025 operating loss 為 \$20.92B。	-\$20.92B	USD / 年	2025	estimate	Quartz / FT 核驗；媒體報導外洩審計文件；URL	"The operating loss for 2025 was \$20.92 billion."	中	估值需要以多年融資能力與損失收斂路徑支撐，而非只看高成長。
O09	OpenAI	成本	Greg Brockman 表示 OpenAI 2026 年將花 \$50B 在 computing。	\$50B	USD / 年	2026	company_comment	Bloomberg 訪談報導；管理層說法；URL	"Spend \$50 Billion on Computing in 2026"	中	compute spend 可能同時進入 COGS、R&D 或資產 / 預付款；應追蹤現金流分類與承諾到期表。

ID	公司	主題	Datapoint / Claim	數值或說法	單位	日期	類型	來源	原文證據	信心	投資含義
C01	TSMC	Foundry 比較	TSMC 2Q26 gross margin 為 67.7% ◦	67.7%	毛利率	2026Q2	fact	TSMC 2Q26 presentation；公司財報；URL	"Gross Margin 67.7%"	高	成熟領先 foundry 即使重資產仍可達高毛利；前提是製程領先與高利用率。
C02	TSMC	Foundry 比較	TSMC 將 2026 capital budget 上調至 \$60-64B ◦	\$60-64B	USD / 年	2026-07-16	company_comment	TSMC 2Q26 transcript；公司指引；URL	"raise our full-year 2026 capital budget ... USD60 billion and USD64 billion"	高	AI lab 的年度 compute 支出已接近單一領先 foundry 的 capex 規模，但多為租賃 / 服務承諾。
C03	Microsoft	雲端比較	FY26 Microsoft Cloud gross margin 為 66% ◦	66%	毛利率	FY2026	fact	Microsoft 10-K；SEC；URL	"Microsoft Cloud gross margin percentage decreased to 66%"	高	雲端以共用資產池與多產品攤銷，毛利顯著高於 OpenAI 2025 推算值。
C04	ServiceNow	AI 軟體比較	FY2025 subscription gross margin 為 80% ◦	80%	毛利率	2025	fact	ServiceNow 10-K；SEC；URL	"Subscription ... 80%"	高	成熟軟體毛利是 AI lab 的上限參照，不是目前基準；推論使用量使成本更變動。
C05	CoreWeave	雲端比較	2025 年 committed contracts 占收入超過 98% ◦	>98%	收入占比	2025	fact	CoreWeave 10-K；SEC；URL	"committed contracts accounted for over 98% ... of our revenue"	高	長約提升供應商融資能力；OpenAI / Anthropic 則承接 take-or-pay 與過度採購風險。
C06	CoreWeave	雲端比較	2025 revenue 為 \$5.1B，年末 RPO 為 \$60.7B ◦	\$5.1B；\$60.7B	USD	2025-12-31	fact	CoreWeave 10-K；SEC；URL	"revenue was \$5.1 billion ... \$60.7 billion of remaining performance obligations"	高	AI compute 供應商的 backlog / revenue 比率很高，突顯基礎設施鏈依賴少數 lab 的信用。
C07	CoreWeave	成本結構	COGS 主要含資料中心租金、電力與折舊攤銷 ◦	rent / power / D&A	成本項	2025	fact	CoreWeave 10-K；SEC；URL	"rent, utilities including power, and depreciation and amortization"	高	AI lab 以租用方式把 capex 轉為長期服務成本，經濟負擔沒有消失。
S01	Anthropic	ASIC	Anthropic 公開採用 AWS Trainium、Google TPU 與 NVIDIA GPU 三種平台 ◦	3 種平台	平台數	2026-02-12	fact	Anthropic Series G；公司披露；URL	"AWS Trainium, Google TPUs, and NVIDIA GPUs"	高	降低單一晶片供應商風險，但增加軟體移植、kernel 優化與排程複雜度。
S02	Anthropic	ASIC	Anthropic 與 AWS Annapurna Labs 共同優化未來 Trainium，含 low-level kernels 與 Neuron stack ◦	共同優化	策略	2024-11-22	fact	Anthropic / AWS；公司聯合披露；URL	"writing low-level kernels ... contributing to the AWS Neuron software stack"	高	接近客製 ASIC co-design，但不承擔完整晶片所有權與量產資本。
S03	OpenAI	ASIC	OpenAI 與 Broadcom 合作部署 10GW OpenAI-designed accelerators，目標 2029 年底完成 ◦	10GW	容量	2025-10-13	company_comment	Broadcom / OpenAI press release；公司前瞻；URL	"deploy 10 gigawatts of OpenAI-designed AI accelerators"	高	若成功，可降低每 token 成本與 NVIDIA 議價依賴；量產良率、HBM / 封裝仍是外部瓶頸。
S04	OpenAI	ASIC	首顆 Jalapeño 為 OpenAI 推論處理器；Broadcom 稱九個月完成設計至 tape-out ◦	9 個月	開發期	2026-06-24	company_comment	Broadcom / OpenAI press release；公司說法；URL	"co-developed ... to manufacturing tape-out in just nine months"	中	證明自研路線進入 silicon 階段；最終 TCO 與量產規模尚未公開。
S05	OpenAI	ASIC	Jalapeño 初步測試僅稱 performance per watt 顯著優於當前 state-of-the-art，最終數據仍在測量 ◦	未公布完整最終值	—	2026-06-24	company_comment	Broadcom / OpenAI press release；公司前瞻；URL	"still measuring final performance"	中	不得把早期 benchmark 當已量產成本優勢；需等實際 tokens/W、良率與部署占比。
K01	Anthropic	Compute 合約	Anthropic 對 AWS 承諾逾 \$100B、十年、最高 5GW ◦	>\$100B；10 年；5GW	USD / 期間 / 容量	2026-04-20	fact	Anthropic / Amazon；公司披露；URL	"committing more than \$100 billion over the next ten years ... up to 5GW"	高	鎖定 Trainium 供給與成本，但形成長期需求 / 技術綁定。
K02	Anthropic	Compute 合約	Google Cloud 擴充協議價值為 tens of billions，預計 2026 上線遠超 1GW ◦	數百億；>1GW	USD；容量	2025-10	fact	Anthropic；公司披露；URL	"worth tens of billions ... well over a gigawatt ... in 2026"	高	多雲採購增加彈性；同時形成多個最低承諾與資本需求。
K03	Anthropic	Compute 合約	Anthropic 承諾採購 \$30B Azure compute，並另約最高 1GW ◦	\$30B；≤1GW	USD；容量	2025-11-18	fact	Microsoft / NVIDIA / Anthropic；公司聯合披露；URL	"purchase \$30 billion of Azure compute capacity ... up to one gigawatt"	高	把 Claude 帶入 Microsoft enterprise distribution，同時深化 NVIDIA 路線。
K04	Anthropic	Capex 安排	Anthropic 宣布與 Fluidstack 在美國建置 \$50B 專用資料中心 ◦	\$50B	USD 投資計畫	2025-11-12	company_comment	Anthropic；公司公告；URL	"announcing a \$50 billion investment in American computing infrastructure"	高	比純 API 公司更重資本；需在 IPO 文件確認誰持有資產、租賃期限與資金來源。
K05	Anthropic	Compute 合約	SpaceX 合作已確認，但官方頁未披露合約金額與期限 ◦	未找到可靠公開金額	—	2026-05-06	fact	Anthropic；公司公告；URL	"a partnership with SpaceX ... substantially increase our compute capacity"	高	資料庫中的 \$26B/year 與可取消條款未獲官方交叉驗證，應視為 rumor，不納入承諾總額。

ID	公司	主題	Datapoint / Claim	數值或說法	單位	日期	類型	來源	原文證據	信心	投資含義
K06	OpenAI	Compute 合約	Stargate 初始目標是在四年投資 \$500B，立即部署 \$100B。	\$500B / 4 年；\$100B 初始	USD	2025-01-21	company_comment	OpenAI / SoftBank；公司計畫；URL	"intends to invest \$500 billion over the next four years"	高	是基礎設施平台意向，不等於 OpenAI 已入帳 capex；股權與專案融資分擔至關重要。
K07	OpenAI	Compute 合約	Oracle / OpenAI 合作規模超過 \$300B，五年，對應最高 4.5GW 追加容量。	>\$300B；5 年；4.5GW	USD / 容量	2025-09-23	fact	OpenAI；公司披露；URL	"partnership that exceeds \$300 billion ... over the next five years"	高	單一供應商集中度與交付風險極大；但可將建置資產與融資留在 Oracle / 專案層。
K08	OpenAI	Compute 合約	OpenAI 對 Azure 新增 \$250B 採購承諾，Microsoft 不再享有算力 ROFR。	\$250B	USD	2025-10-28	fact	Microsoft 10-Q；SEC；URL	"contracted to purchase an incremental \$250 billion of Azure services"	高	保持 Azure 深度綁定又開放多雲；承諾期限與最低採購曲線需在 OpenAI S-1 揭露。
K09	OpenAI	Compute 合約	CoreWeave 披露 OpenAI 承諾最高 \$11.9B 至 2030-10。	≤\$11.9B	USD	2025-03 至 2030-10	fact	CoreWeave 10-K；SEC交易對手；URL	"committed to pay us up to approximately \$11.9 billion"	高	neo-cloud 提供部署速度與供應多元化；OpenAI 承擔供應商信用與 GPU 世代折舊風險。
K10	OpenAI	Compute 合約	Cerebras SEC 文件披露 OpenAI 750MW 合作價值超過 \$20B。	>\$20B；750MW	USD；容量	2026-01	fact	Cerebras S-1/A；SEC交易對手；URL	"multi-year deal ... valued at more than \$20 billion"	高	以 wafer-scale 專用推論降低 latency；鎖定新架構也增加平台與量產風險。
K11	OpenAI	Compute 合約	AMD / OpenAI 確認多世代 6GW GPU 協議，首 1GW MI450 於 2H26 開始。	6GW；首 1GW	容量	2025-10-06	fact	AMD 8-K；SEC；URL	"Initial 1 gigawatt ... starting in 2H 2026"	高	GPU 第二來源改善供應與議價；另有認股權使交易含 circular financing 成分。
K12	OpenAI	Compute 合約	NVIDIA / OpenAI LOI 規劃至少 10GW；NVIDIA 擬按部署進度投資最高 \$100B。	≥10GW；≤\$100B 投資	容量；USD	2025-09-22	company_comment	NVIDIA；LOI / 公司前瞻；URL	"intends to invest up to \$100 billion ... as each gigawatt is deployed"	中	供應商融資降低近期資金缺口，但造成估值、需求與供應商收入的循環關係。
K13	OpenAI	Capex 安排	SB Energy 初始租約為 1.2GW，OpenAI 與 SoftBank 同時投資 \$1B 於 SB Energy。	1.2GW；\$1B	容量；USD	2026-01-09	fact	OpenAI / SB Energy；公司披露；URL	"OpenAI signs 1.2 GW data center lease"	高	以租賃與能源開發商承攬資產層建置；OpenAI 仍負長期租金與利用率風險。
V01	Anthropic	估值	Series H 融資 \$65B，post-money valuation \$965B。	\$65B；\$965B	USD	2026-05-28	fact	Anthropic Series H；公司披露；URL	"raised \$65 billion ... valuing the company at \$965 billion post-money"	高	估值已接近兆美元，未來回報高度依賴 IPO 流動性、持續超高速增长與毛利證明。
V02	OpenAI	估值	最新公司公告融資承諾 \$122B，post-money valuation \$852B。	\$122B；\$852B	USD	2026-03-31	fact	OpenAI funding announcement；公司披露；URL	"\$122 billion in committed capital at a post money valuation of \$852 billion"	高	資本緩衝厚，但 committed capital 的到款條件、稀釋與優先權需核對。
V03	Anthropic	估值	以同期 \$965B valuation / \$47B run-rate revenue 推算約 20.5×	20.5×	valuation / run-rate revenue	2026-05	inference	Anthropic Series H；公司數字 + 研究計算；URL	"run-rate revenue crossed \$47 billion"	高	需要長期維持高成長、接近軟體毛利且壓低資本成本，才可支撐此倍數。
V04	OpenAI	估值	以 \$852B valuation / 8 月媒體報導 \$40B run-rate revenue 推算約 21.3×	約 21.3×	valuation / run-rate revenue	2026-08	inference	OpenAI公司估值 + Axios轉述內部訊息；URL	"OpenAI's latest revenue run rate hit \$40 billion"	中	分子與分母日期不完全一致；僅作敏感度，不應視為正式交易倍數。
V05	Anthropic	估值	2026-02 至 05，估值增約 154%，官方 run-rate revenue 由 \$14B 至 >\$47B。	估值 2.54×；run-rate >3.36×	倍數	2026-02-12 至 05-28	inference	Anthropic Series G/H；公司數字 + 算術；URL	"run-rate revenue is \$14 billion"	高	短期倍數反而壓縮；若成長可持續，Series H 定價不必然比 Series G 更昂貴。
V06	OpenAI	治理 / 估值	Microsoft 持 OpenAI PBC 約 27%，雙方 revenue share 延續，且 Microsoft 仍有部分 IP / API 權利。	約 27%	股權	2025-10-28	fact	Microsoft 10-Q / 官方協議摘要；SEC；URL	"representing roughly 27 percent on an as-converted diluted basis"	高	新投資人需看 revenue share、IP 權利與 AGI 條款對可分配經濟價值的影響。
V07	Anthropic	治理 / 估值	Amazon 在 2026 compute 協議中再投 \$5B，未來可再投最高 \$20B；此前已投 \$8B。	\$5B 現在；≤\$20B 未來；既有 \$8B	USD	2026-04-20	fact	Anthropic / Amazon；公司披露；URL	"Amazon is investing \$5 billion ... up to an additional \$20 billion"	高	供應商、分銷商與股東角色重疊；可能降低資金成本，也增加經濟依賴與關係人定價疑慮。
V08	兩者	估值風險	公開合作金額不可直接相加為負債或 capex，因包含 LOI、投資、租賃、服務採購與專案融資。	不可直接加總	方法論	2026-09-01	inference	本報告綜合 K01-K13；研究推論	"intends to invest" / "committed to purchase"	高	盡調須取得 obligation waterfall：已簽 / 可取消、最低採購、預付款、抵押、到款條件與重疊容量。

註：公司披露的 run-rate revenue 仍是公司定義口徑；媒體取得內部 / 外洩文件者標為 estimate。所有英文摘錄均控制在 25 個英文單字以內。

1. 已確認的事實

Anthropic

- Series H：\$65B 融資、\$965B post-money；公司同月披露 run-rate revenue >\$47B。
- 算力採多元平台：Trainium、TPU、NVIDIA GPU。
- 已公開的主要承諾包括 AWS >\$100B / 10 年 / 最高 5GW、Azure \$30B / 最高 1GW、Google Cloud 數百億 / >1GW、Fluidstack \$50B。
- 正式 GAAP gross margin 未找到可靠公開資料。

OpenAI

- 最新公告融資 \$122B committed capital、\$852B post-money。
- Azure incremental commitment \$250B；Oracle 合作 >\$300B / 5 年；CoreWeave ≤\$11.9B；Cerebras >\$20B / 750MW。
- 與 Broadcom 的自研 ASIC 已走到首顆推論處理器；10GW 計畫仍屬前瞻部署目標。
- Microsoft 持 PBC 約 27%、仍有 revenue share 與特定 IP / API 權利。

2. 仍屬估算或傳聞的內容

項目	目前證據	應如何使用
OpenAI 2025 revenue / COGS / R&D / loss	外洩 audited financials、Financial Times 與多家媒體核驗；OpenAI 未公開發布。	可作中信心 estimate；待 S-1 對照會計口徑與關係人交易。
Anthropic 2Q26 adjusted operating profit	Bloomberg 見內部文件；公司未發布完整損益表。	視為中信心 estimate；需核對 SBC、訓練支出資本化 / 調整項。
Anthropic SpaceX 合約 \$26B/year、90 天可取消	Stock LLM Wiki 指向單一 SemiAnalysis 報告；官方只確認合作、未披露金額。	列為 rumor、未納入估值或承諾總額。
OpenAI / NVIDIA \$100B	正式新聞稿是 letter of intent、且投資隨每 GW 部署。	不是一次性已到款資金；估值模型應按里程碑折現。

3. Anthropic vs OpenAI 的差異

維度	Anthropic	OpenAI	投資判讀
主要商業動能	企業、開發者、Claude Code、usage-based API	消費訂閱 + 企業/API + 更廣產品面	Anthropic mix 較利於 usage monetization；OpenAI TAM 與分銷面更廣。
晶片策略	不以完全自有晶片為主；深度共同優化 Trainium、並採 TPU、NVIDIA	自研 Broadcom ASIC + NVIDIA / AMD / Cerebras 多架構	Anthropic 降低 tape-out 風險；OpenAI 保留更大長期 TCO 上行。
資本安排	多雲採購 + Fluidstack 專用設施 + 供應商股權投資	Stargate / Oracle / Azure / neo-cloud / 晶片商、且供應商融資更明顯	OpenAI 的容量選擇權更多，但合約 waterfall 更複雜。
財務透明度	官方 run-rate 多；完整 GAAP 數據少	官方 run-rate 少；外洩 audited figures 較完整但未正式發布	兩者都需等 S-1；不可用不同口徑直接排名。

4. Token economics 與毛利的關鍵驅動因素

Revenue per workload

- input / output token mix
- reasoning 與 agent tool-loop 長度
- API、seat subscription、enterprise committed spend
- 模型路由與 premium tier mix

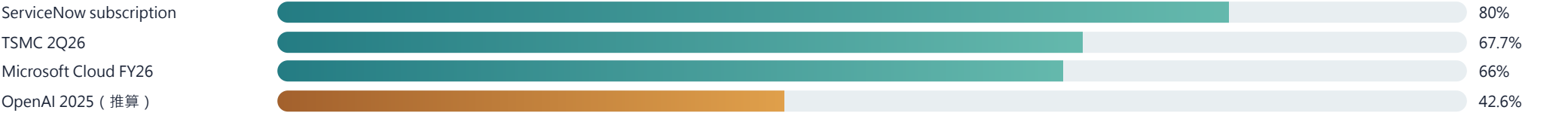
Serving cost per token

- tokens/sec、tokens/W、batching
- cache hit rate / KV cache 占用
- HBM、network、power、cooling
- GPU / ASIC 利用率與故障冗餘

Below gross profit

- pre-training / post-training compute
- research talent 與 SBC
- 安全、資料授權、銷售與支持
- 長約預付款、融資利息、資產減損

公開毛利參照



注意：不同公司會計口徑、期間與產品 mix 不同；此圖只用於結構參照。Anthropic 未有可靠公開 GAAP gross margin。

5. 最值得追蹤的後續 datapoints

優先度	Datapoint	為何重要	可接受的驗證來源
1	S-1：GAAP revenue、gross profit、R&D、operating cash flow、SBC	把 run-rate 與 adjusted narrative 轉成可比財務。	SEC S-1 / S-1A、審計報告
2	Compute obligation maturity ladder	判斷每年最低現金需求、取消權與重疊容量。	S-1 commitments、供應商 10-K exhibits
3	Inference COGS per 1M effective tokens	驗證定價下降能否被 tokens/W、cache、batching 抵銷。	公司 KPI、審計毛利 bridge、硬體 telemetry
4	API / subscription / enterprise mix 與 net revenue retention	區分高使用量負毛利訂閱與可擴張企業 API。	S-1 segment / cohort disclosure
5	Jalapeño 實際量產占比、yield、tokens/W、HBM 與封裝配置	判斷 OpenAI 自研 ASIC 是否真的創造成本護城河。	OpenAI、Broadcom、TSMC、封裝供應鏈公告
6	Trainium / TPU / NVIDIA 在 Claude workload 的分配	驗證 Anthropic 多平台策略是韌性還是低效率。	Anthropic / AWS / Google telemetry 或客戶案例
7	關係人 revenue share 與供應商投資條款	確認 reported revenue 有多少需返還、估值有多少由供應商融資支撐。	S-1 related-party notes、投資協議

6. 可能推翻目前投資 thesis 的反證

推翻 Anthropic 相對優勢

- S-1 顯示 GAAP gross margin 明顯低於 50%，且 adjusted profit 主要靠排除 SBC / 訓練成本。
- Claude Code 成長在 2-3 季內快速降速，enterprise expansion / NRR 不足。
- 多雲 porting 與容量閒置使 effective compute cost 高於單一平台。
- AWS / Google / Azure 合約含高額不可取消最低採購，超過可見營收。

推翻 OpenAI 長期作業槓桿

- Jalapeño 量產延誤或實際 tokens/W 未優於外購方案。
- 年度 compute spend 持續快於 revenue，2025 約 42.6% 推算毛利無改善。
- Oracle / Azure / GPU 合約出現 take-or-pay 減損或 refinancing 壓力。
- Microsoft revenue share / IP 權利讓 OpenAI 的可分配經濟價值顯著低於 headline valuation。

估值框架與盡調結論

框架	Anthropic	OpenAI	pre-IPO 投資者要求的安全邊際
Headline multiple	約 20.5× 同期官方 run-rate	約 21.3× 不同日期 run-rate；較低信心	至少以 3 個分母重算：LTM GAAP revenue、net revenue after rev-share、forward contracted revenue。
Margin normalization	勿用「API mix」直接假設 70-80%	2025 推算約 42.6%、訓練 R&D 另計	以 steady-state inference GM、維持前沿所需 R&D / revenue、SBC 稀釋三層建模。
Capital intensity	多雲承諾 + Fluidstack 專用設施	Stargate + 多個超大型採購 / 租約	使用 lease-adjusted FCF 與 committed-spend-adjusted enterprise value。
Optionality	企業 coding 與跨雲分銷	消費入口、自研 ASIC、廣泛硬體供應	只為已觀察到的 cohort economics 付費；未量產 ASIC、廣告與 AGI optionality 應大幅折現。
主要結論	近期財務質量看似較佳，但公開會計資料不足。	規模與基礎設施選擇權較大，但資本與治理結構更複雜。	在 S-1 前，較適合以條件式投資、估值保護、資訊權與 milestone tranches，而非單一 headline price。

方法與資料限制

- 研究截止日依使用者 placeholder 解讀為 2026-09-01。
 - Fact 指公開可直接驗證；company_comment 指管理層績效 / 未來說法；estimate 指媒體或研究機構估算；rumor 指未確認消息；inference 指本報告計算或判斷。
- 私有公司 run-rate、adjusted profit 與內部文件並非 GAAP 定期申報；明確標示資料來源層級。
 - 不同來源衝突時不平均；例如 Anthropic SpaceX 金額只見單一資料庫來源，官方未披露，故保留為 rumor。
 - 本報告不是投資建議；pre-IPO 證券可能有鎖定期、轉讓限制、優先權與資訊不對稱。

主要來源索引

- Anthropic Series H (2026-05-28)
 - Anthropic Series G (2026-02-12)
 - Anthropic–Amazon compute (2026-04-20)
 - Anthropic–Google Cloud
 - Microsoft–NVIDIA–Anthropic (2025-11-18)
 - Anthropic–Fluidstack (2025-11-12)
 - OpenAI funding (2026-03-31)
 - Stargate (2025-01-21)
 - OpenAI–Oracle Stargate (2025-09-23)
- Microsoft 10-Q / OpenAI \$250B Azure commitment
 - CoreWeave 2025 10-K
 - Cerebras S-1/A
 - AMD 8-K / OpenAI 6GW
 - NVIDIA–OpenAI LOI
 - Broadcom–OpenAI Jalapeño (2026-06-24)
 - Ars / FT on OpenAI audited figures
 - Bloomberg on Anthropic 2Q26 internal documents
 - Stock LLM Wiki：Anthropic (未)、OpenAI (未)、技術_AI推論與ASIC平台、報告_SemiAnalysis_Anthropic_IPO財務分析_20260708。